Human brain-to-brain synchrony in a naturalistic setting: an fMRI study on observational learning

Kathrin Kostorz

Dissertation der
Graduate School of Systemic Neurosciences
der Ludwig-Maximilians-Universität München
29.05.2019

Human brain-to-brain synchrony in a naturalistic setting: an fMRI study on observational learning

Kathrin Kostorz

Graduate School of
Systemic Neurosciences
LMU Munich

Dissertation der
Graduate School of Systemic Neurosciences
der Ludwig-Maximilians-Universität München

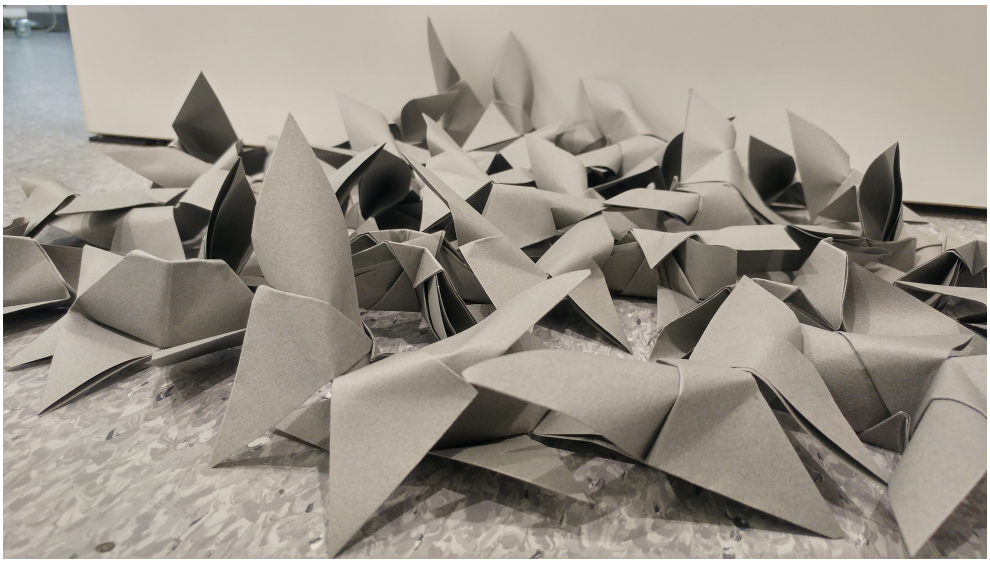
München 2019

Supervisor & First Reviewer: Prof. Dr. Stefan Glasauer
Second Reviewer: Dr. Virginia Flanagin
Third Reviewer: Dr. Merle Fairhurst

Date of Submission: 2019-05-29
Date of Defense: 2019-09-30

We are all explorers, driven to know what's over the horizon,
what's beyond our own shores.
And yet, the more I've experienced, the more I've learned that
no matter how far we travel or how fast we get there, the most
profound discoveries are not necessarily beyond that next star.
They're within us; woven into the threads that bind us, all of us,
to each other.

Captain Jonathan Archer
commander Enterprise NX-01



Instead of one thousand cranes, we offer one thousand butterflies.

Abstract

In today's society we are highly connected through media. Other people's knowledge and skills are publicly available: Anybody can distribute a video in which they are instructing or teaching a task, and anybody interested can access these videos. By observing these videos, we can gain insights into the processes demonstrated or learn to reproduce the task by imitating the instructor of the video.
In this thesis, I would like to investigate these processes of action learning. To do this, I have chosen the production of a handicraft: the folding of a paper figure, also known as origami.

Using functional magnetic resonance imaging (fMRI), I recorded the brain activity of an instructor who was blindfolded and created an origami while being videotaped. I then recorded the brain activity of several observers who watched these videos with the task of reproducing the origami right afterwards. As a control condition in which learning was not required, I utilized a video where the instructor produced partly repetitive paper folds. Here, the observers were required to count how many folds were made.
This thesis examines the relationship between the instructor and the observers on the one hand and between the observers themselves on the other hand. This experiment was kept as naturalistic as possible with no constraints on the participants. This leads to an open, data-driven design. To account for this, I investigate the neural relationship between different agents by determining the synchrony of corresponding brain areas. I utilize the intersubject correlation method using a jackknife routine for instructor-observer pairs, and a principal component analysis for the observers between themselves. Potential time shifts between the agents are determined via a lag bootstrap routine.

In this thesis, I demonstrate a consistent synchrony in areas of the action observation execution network (AOEN) across different agents and across a diverse range of cognitive tasks. Also, better subjects tend to show higher synchronization to the instructor in the

premotor cortex. On the other hand, changes in synchrony are located largely outside classical AOEN areas, possibly reflecting changes in common processes of attention and spatial working memory. In particular, this thesis demonstrates vast changes in synchrony between different cognitive tasks and learning stages in the cerebellum. These changes are located largely outside the classical hand sensori-motor areas. This is considerably intriguing, given that to this day, the cerebellum is a sparsely researched area in the fMRI literature, and as a 'motor area' neglected in cognition research.

Contents

Abstract **vii**

Contents **ix**

List of Abbreviations **xiii**

1 Introduction **1**

1.1 Naturalistic studies in an unnatural environment - stories, movies and human interaction inside a Magnetic Resonance Imaging machine 1

1.1.1 Overview 1

1.1.2 Let's watch some movies! 2

1.1.3 Social interaction inside a big, loud and isolated machine 3

1.1.4 The challenges and thrills of analyzing an open design - finding temporal similarities 5

1.2 I see what you are doing - action processing and observational learning in the brain 8

1.2.1 Actions and observations 8

1.2.2 Mirror neurons 8

1.2.3 The action observation network 10

1.2.4 The action observation execution network 11

1.2.5 Involvement and modulation of the action observation execution network in naturalistic contexts of experience 13

1.3 Motivation for this thesis 18

2 Synchronization between instructor and observer when learning a complex bimanual skill **21**

2.1 Introduction 22

2.2 Materials and methods 26

2.2.1 Participants . . . 26
2.2.2 Stimuli . . . 26
2.2.3 Experimental MRI setup . . . 27
2.2.4 Experimental procedure . . . 28
2.2.5 fMRI acquisition and preprocessing . . . 31
2.2.6 Data analysis . . . 32
2.2.7 Behavioral rating of the observer performance . . . 36
2.2.8 ROI analysis . . . 37
2.3 Results . . . 38
2.3.1 Behavioral results . . . 38
2.3.2 Synchronization between observers and instructor . . . 40
2.3.3 Instructor-observer correlations during observational learning . . . 44
2.3.4 Instructor-observer correlations across learning . . . 45
2.3.5 Temporal shifts . . . 49
2.4 Discussion . . . 51
2.4.1 Overall observer-instructor synchronization . . . 51
2.4.2 Memorizing/encoding vs. identifying actions . . . 53
2.4.3 Synchrony reduces during repeated viewing . . . 55
2.4.4 Better reproduction is associated with higher instructor-observer correlation in the left ventral premotor cortex . . . 56
2.4.5 Exploring the temporal relations . . . 58
2.4.6 Cerebellar contributions . . . 59
2.4.7 Limitations and outlook . . . 60
2.4.8 Conclusion . . . 61
2.5 Supplementary material . . . 62

3 Synchrony between observers during observation and imitation learning of origami 65
3.1 Introduction . . . 66
3.2 Materials and methods . . . 69
3.2.1 Experimental setup and procedure . . . 69
3.2.2 Data analysis . . . 71
3.3 Results . . . 75
3.3.1 Overall similarity during action observation . . . 75
3.3.2 Commonness between sessions . . . 76
3.3.3 Difference between sessions . . . 79
3.3.4 Relation of the common component to the optic flow of the videos 83
3.4 Discussion . . . 88
3.4.1 Involvement of the AOEN and beyond . . . 88

3.4.2 Optic flow as a putative proxy for hand movement 91
3.4.3 M1 involvement 93
3.4.4 Conclusion and outlook 94
3.5 Supplementary material 96

4 General discussion **99**
4.1 Discussion of the main findings 99
4.1.1 Overview 99
4.1.2 Synchrony of temporal activity 100
4.1.3 What areas belong to the AOEN? 101
4.1.4 Reflections about behavior and performance 103
4.2 Further research ideas 105
4.2.1 Expanding knowledge using this data set 105
4.2.2 Further research ideas for follow-ups 107
4.3 Relevance of this thesis 109
4.3.1 Direct contributions 109
4.3.2 Outlook 111
4.3.3 Conclusion 113

Bibliography **115**

A Appendix **131**
A.1 Acknowledgments 132
A.2 Eidesstattliche Versicherung/Affidavit 134
A.3 Author contributions 135

List of Abbreviations

fMRI	functional magnetic resonance imaging
BOLD	blood oxygen level dependent
EEG	electroencephalography
fNIRS	functional near-infrared spectroscopy
AON	action observation network
AOEN	action observation execution network
MNS	mirror neuron system
PMC	premotor cortex
vPMC	ventral premotor cortex
dPMC	dorsal premotor cortex
SMA	supplementary motor area
DLPFC	dorsolateral prefrontal cortex
IFG	inferior frontal gyrus
IPL	inferior parietal lobule
SPL	superior parietal lobule
M1	primary motor area
LOTC	lateral occipito-temporal cortex
TPJ	temporo-parietal junction
aIPS	anterior intraparietal sulcus
ISC	intersubject correlation
PCA	principal component analysis

1

Introduction

1.1 Naturalistic studies in an unnatural environment - stories, movies and human interaction inside a Magnetic Resonance Imaging machine

1.1.1 Overview

Naturalistic experiments inside a functional magnetic resonance imaging (fMRI) machine - that seems to be a contradiction in terms, given that the subjects have to lie inside a loud -and for some people scary- machine. Spatially constrained, the words 'do not move your head' still ringing in your head, and receiving stimuli in a supine position that most people only adopt to fall asleep - these aren't the ingredients for a naturalistic experiment. And yet, these are the constraints for performing any kind of fMRI experiment. So why should we use functional MRI? Electroencephalography (EEG) or functional near-infrared spectroscopy (fNIRS) offer a certain freedom of position and movement for the subjects and a great temporal resolution for brain imaging. However, they come with a major drawback: only cortical structures close to the skull can be imaged and that at a spatial resolution that is much lower than what fMRI can achieve. Therefore, fMRI is the go-to method in non-invasive human neuroscience if fine spatial structures or deeper cerebral structures like the basal ganglia or the cerebellum need to be imaged. Naturalistic experiments that require fine cerebellar imaging can only be done inside an MRI machine.

Naturalistic experiments in this case mean that the presented stimuli are more related to the everyday experience of the subjects, such as photographs, movies, real-life sounds, etc. These are in contrast to classical psychological experiments that present very reduced but therefore highly controlled stimuli like dots, bars, arrows, etc. Using more naturalistic input we lose some of the control over our stimuli (e.g. the image of a tree is made of several different stimuli like colors or texture or aspects like leaves), but we gain

experiences that are closer to everyday life and the environment we live in.
Using functional MRI, the typical studies involve all kinds of photography (natural scene perception, see e.g. Walther et al., 2009 or Stansbury et al., 2013). But several studies have gone even further to human recreational everyday life using parts of feature films with or without sound (Hasson et al., 2004, 2008a,b), music (e.g. Abrams et al., 2013), or recorded stories which are told at storytelling events (Lerner et al., 2011) or even inside the MRI machine (Stephens et al., 2010; Zadbood et al., 2017).

1.1.2 Let's watch some movies!

The first study which presented subjects parts of a feature film (in this case 'The Good, the Bad and the Ugly') inside the MRI was done by Hasson et al. (2004). Looking at the MRI images, they could show enhanced similarity of the recorded neural activation between subjects when people watched the movies. In darkness, the extent of the between-subjects correlated surface was low. Moreover, they also demonstrated peaks of activity in certain areas when the subjects observed the corresponding stimuli (e.g. the fusiform face areas showed peaks for scenes with face close-ups). This was an important step for fMRI research given that before mainly strongly controlled stimuli were presented, but not a totally uncontrolled visual stream like a movie. Moreover, previously, the typical approach was to contrast different stimuli presented in a defined order. Here, not only is the movie itself uncontrolled, but also a control condition with which to contrast the between-subjects correlation is apparently missing.
Here, the fact that there is a similarity in a certain voxel/region is thought to implicate a common processing of the stimuli in this voxel/region. Otherwise, there would not be any similarity of the signal between the subjects. Needless to say, a new analysis different from the traditional GLM approach had to be employed. Hasson et al. (2004) have introduced the intersubject correlation. This is essentially a Pearson correlation of the voxelwise time course between subjects, which is used as a measure of intersubject similarity or synchrony (see section 1.1.4 for the mathematical formalities). Several follow-ups and new studies were conducted not only to broaden the handling with these stimuli (movies and audio stories) but also to establish the intersubject correlation method. Diverse studies tried to gain insights into different aspects of this naturalistic processing by controlling the presented movie stimuli to some extent: this ranges from investigating the affective processing (by showing movie clips of different valence and arousal ratings, see Nummenmaa et al., 2012) to temporal processing (by scrambling parts of the movie on different timescales, e.g. Hasson et al., 2008b) and perspective-taking or theory of mind (by showing the same movie clips but with different additional tasks of mentalizing/perspective-taking, see Lahnakoski et al., 2014).
It should be noted, however, that while the fact that feature films are naturalistic in the

sense that they depict natural stimuli and that they play a great role in our culture, they are not necessarily naturalistic in terms of the input we gain in our daily living. In contrast to our everyday stimuli in the outside world, feature films are highly edited audio-visual input. Herbec et al. (2015) have compared a professionally edited movie of a solo dance with an unedited one. (Here, the same dance stimulus was recorded with several cameras. One movie was just the video stream of one long shot, the other movie combined cuts of other camera recordings like close-ups mounted while audio and content stayed the same). They found higher intersubject correlation for the edited video than for the unedited one. While the exact reason for this remains speculative (one possible explanation is overall higher attention due to attention captures, see e.g. Cohen et al., 2015), it is important to note that there is difference between feature films and everyday stimuli that might be due to the professional editing.

1.1.3 Social interaction inside a big, loud and isolated machine

Human social neuroscience using fMRI has traditionally focused on showing participants social stimuli like emotional faces, different gazes or depicted interactions between people or of people with objects (Adolphs, 2003). However, a great part of human sociality is the relation to each other or the interaction with each other. This importance has led authors like Schilbach (2016) to claim that psychiatric disorders are also disorders of social interaction and should be researched as such. Consequently, social interaction plays an enormously meaningful part in our human life.
Figure 1.1 shows different types of human relation or interaction. Using fMRI, less has been done to actually investigate the interaction of the participants (instead of observation of interactions), and even less has been done to investigate both interacting parties

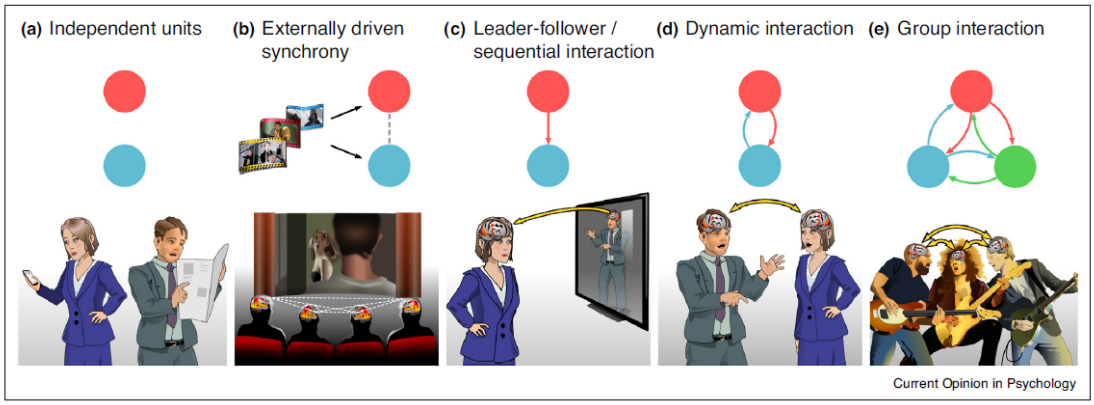


Figure 1.1: Different types of interaction and corresponding brain-to-brain synchrony. Reprinted with permission from Nummenmaa et al. (2018), license number 4597620917323.

inside an MRI machine to obtain both parties' neural recordings. On this side, the first so-called hyperscanning study was performed by Montague et al. (2002). They also created the term 'hyperscanning' to describe human interaction studies where both subjects are scanned simultaneously (via synchronized technical devices). Putting two subjects into different MRI machines and having them interact on-line was certainly an extensive and painstaking, but also fascinating research attempt. While the first hyperscanning study was based on behavioral animal experiments ('guess the chosen color and receive a bit of juice if correct; otherwise the other player receives it'), later studies examined a broader set: imitating gaze (e.g. Koike et al., 2016), directing information by gaze (e.g. Bilek et al., 2015) or by moving virtual symbols (Stolk et al., 2014; Koike et al., 2019), or telling an autobiographical event to a friend (Spiegelhalder et al., 2014). By transmitting videos of each others' faces and/or speech in real time, one can obtain a form of direct intimate interaction despite the heavily technical and unnaturalistic environment.
While fascinating, the hyperscanning approach is very equipment-heavy, as two MRI machines not only have to be synchronized, but also available for research at the same time. Luckily, for many questions it is not necessary to put two subjects simultaneously inside an MRI machine. Here, a limitation of fMRI allows for a reduction in complexity: Communication in an MRI machine is unfortunately only possible through media like video and sound recordings. (But see the work of Lee et al. (2012) for the invention of an MRI receiver coil allowing two heads to be next to each other).
The crucial point is online interaction: when a direct reaction of one subject to another, continuous feedback, or a fine-grained interaction like a collaboration is to be investigated (see fig. 1.1 d and e), hyperscanning is needed. When the research question does not necessitate online interaction, one can also scan successively. This can be done by showing the second person in the scanner the recorded 'message' of the first person in the scanner (figure 1.1 c). With this approach, some studies have investigated this unidirectional approach to communication, using gestures or recordings of stories. The researched questions range from sharing emotions with your partner (Anders et al., 2011) to playing charades with him/her (Schippers et al., 2009 and Schippers et al., 2010). It also involves telling and listening to stories (Stephens et al., 2010; Zadbood et al., 2017). These studies typically show a form of relationship between the time series of the interacting parties. However, classical GLM analyses need tasks designed in a certain way (e.g. by predefined blocks). To allow for more naturalistic options, only minor design constraints should be imposed on the interaction parties. To analyze the recordings, methods from time series analyses can be used when the traditional design is lost.

1.1.4 The challenges and thrills of analyzing an open design - finding temporal similarities

Intersubject correlation

Correlations and cross-correlations are a classical and easy but efficient measure to examine the temporal relation and synchronicity of time courses (Kreuz et al., 2007). To relate the time courses of different subjects to each other, they were firstly applied in fMRI literature by Hasson et al. (2004). Classical fMRI data analysis relies on creating a general linear model. It involves creating a model of activation/deactivation according to the experimental design, which is then convolved with the hemodynamic response function, which is thereupon fitted to the recorded data.
Contrary to this approach, Hasson et al. (2004) did not create a design model but rather postulated that brain areas would temporarily synchronize between subjects when subjects were processing the same stimulus. On the other hand, if there were no common processing, the signals would not be correlated between subjects, as is the case in resting state fMRI. This type of analysis was one of the early approaches to the nowadays more common model-free analysis of fMRI data. Figure 1.2 shows a comparison between the different concepts of the general linear model and the intersubject correlation.

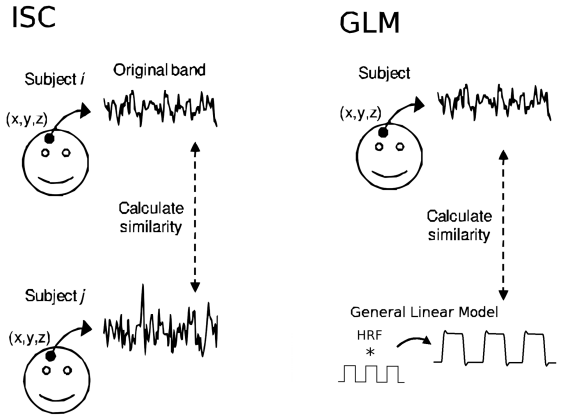


Figure 1.2: Differences between intersubject correlation (ISC) and classical fMRI general linear model (GLM). Reprinted from Pajula et al. (2012) under CC BY license. Image doi: https://doi.org/10.1371/journal.pone.0041196.g001.

The most common measure of correlations between two continuous signals X and Y is the Pearson correlation coefficient

$$r_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \tag{1.1}$$

with cov(X,Y) being the covariance between signals X and Y, and σ_X and σ_Y being the respective standard deviations. For the first manuscript in this thesis (chapter 2), we have essentially calculated the Pearson correlation coefficients between the instructor

and each observer for each voxel. As N instructor-observer pairs can be formed, this resulted in N correlation coefficients per voxel (with N being the number of subjects) and per condition. This distribution of N coefficients was then conventionally assessed using t-tests.
If we want to compare the synchrony between the N subjects, this gets more complicated. In previous papers, typically one combined the subjects into pairs, which results in $N \cdot (N-1)/2$ pairs. Consequently, $N \cdot (N-1)/2$ correlation coefficients have been computed and averaged, as in e.g. Kauppi et al. (2014). Another option is to compute per subject the mean time course of all other $N-1$ subjects, and then correlate the respective two time courses (subject and mean). This results in $N-1$ correlation coefficients, see e.g. Hasson et al. (2008b); Stephens et al. (2010). These two methods are mathematically analogous (Kauppi et al., 2014).

Principal component analysis

To avoid constructing these pairwise coefficients and averaging, we propose to utilize a more expansive procedure which we use in chapter 3. We perform a principal component analysis (PCA) on the raw voxel-wise data set $\mathbf{X}$ with dimensions N×T (N samples/subjects and T observations, with T being the number of time points). This common dimensionality reduction technique (Abdi and Williams, 2010b) can be regarded as a generalization of the 2D correlation approach as it computes the same idea: how much variance can be explained by the 'commonness' of the variables.
Essentially, a principal component analysis is an eigenvalue problem of the covariance matrix $\mathrm{cov}(\mathbf{X})$ of the data.

$$\mathbf{W}^T \mathrm{cov}(\mathbf{X}) \mathbf{W} = \mathbf{D} \tag{1.2}$$

with $\mathbf{D}$ being the diagonal matrix of the eigenvalues of the covariance matrix, and $\mathbf{W}$ being the matrix of the eigenvectors.
We are looking for a new coordinate system where the non-diagonal elements of the covariance matrix are zero. The eigenvectors of equation 1.2 span the new coordinate system; these are the principal components. Since the covariance matrix is symmetric, the eigenvectors (the components) are orthogonal (if the eigenvalues are different). Therefore, the scores, which are the projections of the observations (in our case the time courses), are (linearly) independent, which means uncorrelated. (Note that there is no consistent naming in the literature: sometimes, the scores are also called 'components'.)
The eigenvalues of the covariance matrix are the 'new' variances alongside the components. The first principal component is defined as the eigenvector that points alongside the largest variance of the data set (the largest eigenvalue).
How are the eigenvalues related to the previously determined average correlation values? When whitening and normalizing the data, the covariance matrix becomes a correlation

matrix. Friedman and Weisberg (1981) showed that if the correlations are positive or only few small negative correlations exist, the first eigenvalue of the correlation matrix is closely approximated by $\lambda_1 \approx 1 + (N-1) \cdot \bar{r}$ (with λ_1 being the first eigenvalue and $\bar{r}$ the mean correlation). When simulating data, we found a similar good proportional relationship, see also chapter 3 and figure 3.8.

Instead of using the covariance matrix to perform the PCA, a common approach is to compute a singular value decomposition (SVD) of the original data matrix (algorithms for the SVD also computationally preferable). For centered (de-meaned) data $\mathbf{X}_n$ it holds: $\mathrm{cov}(\mathbf{X}_n) = {}^1/_{(N-1)}{\mathbf{X}_n}^T\mathbf{X}_n$. The singular value decomposition of $\mathbf{X}_n$ is: $\mathbf{X}_n = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ with $\mathbf{\Sigma}$ having the singular values on the diagonal. Then:

$$\frac{1}{N-1}{\mathbf{X}_n}^T\mathbf{X}_n = \frac{1}{N-1}(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T) = \frac{1}{N-1}(\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T)(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T) \tag{1.3}$$

$$= \frac{1}{N-1}\mathbf{V}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{V}^T \tag{1.4}$$

This means ${\mathbf{X}_n}^T\mathbf{X}_n$ and $\mathbf{\Sigma}^T\mathbf{\Sigma}$ are similar matrices and therefore have the same eigenvalues. Since $\mathbf{\Sigma}^T\mathbf{\Sigma}$ is a diagonal matrix, for the singular values s_i it holds

$$s_i^2 = (N-1)\lambda_i \tag{1.5}$$

with λ_i being the eigenvalue of $\mathrm{cov}(\mathbf{X}_n)$. This also turns $\mathbf{V}$ into the eigenvectors of ${\mathbf{X}_n}^T\mathbf{X}_n$ and therefore $\mathrm{cov}(\mathbf{X}_n)$.

When simulating data we also observed a very good linear relation between the average correlation value and a separability index S:

$$S_1 = \frac{s_1^2}{\sum_{i=1}^{N} s_i^2} \tag{1.6}$$

This is essentially $1-\alpha$ with α being the degree of inseparability proposed by Depireux et al. (2001). Given the relation in equation 1.5, $S_1 = {}^{\lambda_1}/_{\sum_{i=1}^{N}\lambda_i}$ is the proportion of the variance explained alongside the first component in relation to the total variance.

1.2 I see what you are doing - action processing and observational learning in the brain

1.2.1 Actions and observations

To navigate our social environment, we constantly have to make sense of the actions of others. We have to process and interpret them in order to react or interact accordingly (see e.g. Rizzolatti et al., 2001). We also learn new actions and processes by observing and imitating others, be it skiing, knitting or learning to change the tires of a car. Consequently, observational learning, that is learning actions by observing others, is an integral part of human life, especially for children (Bandura, 1971). The availability of new media offers a new facet, which is becoming increasingly popular: e-learning by observing others. Online video platforms offer the option to find videos where an instructor demonstrates how-to procedures as diverse as skiing, knitting or learning to change tires. In order to learn by observation, one has to first attentively watch the stimulus, with the prerequisite of being motivated for the process. Then this stimulus needs to be processed and stored in memory. During reproduction the stimulus has to be retrieved from memory and then enacted accordingly (Bandura, 2004). From a cognitive neuroscience point of view, during stimulus processing and memory de- and encoding the stimulus has to be mapped onto or related to the observer's own motor repertoire (simulation or embodiment theory, e.g. Jacob and Jeannerod, 2005). The system in the brain that is thought to mediate these processes is nowadays referred to as the action observation execution network (AOEN, Thomas et al., 2018) in humans. Below I will give a quick overview of the scientific history of this cognitive neuroscience research field.

1.2.2 Mirror neurons

In the 1990s, a research team based in Parma made an incidental finding when measuring neuronal discharge in monkeys. Neurons in the primate premotor area F5 not only fired when the monkey performed an action, but also when it observed another agent performing the same or a similar act (Di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 2001; Rizzolatti and Craighero, 2004). Given that the premotor areas had been considered as pure motor planning areas, this finding was very intriguing: Neurons in a motor area spiking without a motor act performed, just by observing another agent's motor actions. The researchers called these neurons 'mirror neurons'. Consequently, mirror neurons are defined as being active when the agent performs an action and when it sees an action.

Subsequently, mirror neurons in the non-human primate cortex were also found in the inferior parietal lobe (Fogassi et al., 2005; Rozzi et al., 2008). These two areas, premotor F5 and inferior parietal lobe (IPL), constitute the core of the classical non-human pri-

mate mirror neuron circuit. It receives input from the superior temporal sulcus (STS) and inferior temporal lobe (motion processing without mirror properties) and is controlled by prefrontal areas Rizzolatti and Sinigaglia (2010).
More recently, neurons with mirror or mirror-like properties have been found in the non-human primate primary motor cortex (Dushanova and Donoghue, 2010; Vigneswaran et al., 2013 and Kraskov et al., 2014). Interestingly, some of these M1 mirror neurons demonstrate suppression behavior, as they are active during action execution but their activity is suppressed during action observation (Kraskov et al., 2014).
The fact that the premotor area F5, which was formerly considered purely a motor area, was now involved in action observation was an intriguing finding to many people. It led to extensive speculations about the purpose of these neurons and areas. Some authors proposed a 'direct-matching' theory (e.g. Rizzolatti et al., 2001), meaning that observed actions are mapped onto an agent's own motor repertoire. This can be embedded into the more comprehensive cognitive theory of mental simulation (Jacob and Jeannerod, 2005). Some researchers have claimed that action understanding (e.g. Rizzolatti et al., 2001, 2002) or goal coding (e.g. Umilta et al., 2008) is induced by this direct matching, but see e.g. (Hickok, 2009, 2013) for a refutation.
Since these findings in non-human primates have been very intriguing, researchers have tried to find similar common neural signatures for action performance and action observations in humans. This motivation has led to a fruitful field of study with manifold publications dating from the 2000s onwards. Common activations for action observation and execution were found in the parietal and premotor cortex, but also beyond (see e.g. Gazzola and Keysers (2009) for a well-controlled fMRI study). For humans, the term 'mirror neuron system' (MNS) was used.
While certainly very interesting results for human mirror neuron research were produced, early research suffered from methodological problems (Turella et al., 2009; Hickok, 2009). Major aspects of these issues can and also have been addressed. For example, one can let the subjects perform the actions blindly to ensure that activations are not solely due to their own visual feedback, meaning their own action observation during action execution. However, other aspects for human mirror neuron research stem from the problematic terminology: The studies utilize mainly typical non-invasive methods like functional MRI or EEG, which makes it debatable when deductions about neurons are made (Turella et al., 2009; Hickok, 2009). The different scopes between measuring a proxy for neural activity like the BOLD signal in humans and the single cell recordings in monkeys render reasoning about human mirror neurons questionable. Sometimes people therefore have referred to it as the 'putative human mirror neuron system' (Keysers, 2009). One approach to overcome this limitation were the studies on repetition suppression such as (Kilner et al., 2009) or (Dinstein et al., 2007), which make the attempt of neuronal

inference explained via fatigue effects. Currently, I only know of one study that uses single-cell recordings to research mirror neurons in humans, i.e. (Mukamel et al., 2010). Here, neurons with mirror properties have been found in different, non-classical MNS areas of the brain: amongst others, the supplementary motor area, hippocampus and anterior cingulate cortex. (On a side note, the methodological problem of self-generated visual feedback is also valid for this study.)

1.2.3 The action observation network

However, when the subject of study is not to research human mirror neurons, but to investigate the human processing of actions and of observations of actions, it is not necessary to infer mirror neurons. In the end, the neural mechanisms of actions and action observations could, but do not have to be mediated by mirror neurons. Similar activations between these two cognitive processes are consistent with the idea that this process is mediated by mirror neurons, but mirror neuron activity is not required to produce similarities between acting and observing actions (see e.g. Gazzola and Keysers, 2009).

If the goal of the study is to investigate human action observation, and not infer neurons, one should use the broader terminology of the 'action observation network' instead of the 'mirror neuron system'. The term 'action observation network' (AON) was coined before (e.g.Calvo-Merino et al., 2006), but has become increasingly popular since 2009 (Grafton, 2009; Cross et al., 2009). Grafton (2009) defines it as the network of brain regions being active when one person observes another person in motion. While this initial definition encompasses only the observation of actions, the AON shares areas with the circuits involved during motor production: the dorsal and ventral premotor cortex are part of the action observation network as well as the IPL, the inferior frontal gyrus (IFG) and the posterior STS (Grafton, 2009). Homologue areas of the classical mirror neuron areas in non-human primates, F5 and IPL, are a subpart of this network. The main advantage of the term AON is that it constitutes a wider definition, more suitable for functional MRI. Importantly, also areas not associated with mirror neurons can be part of this network. Moreover, it would be clearly too reductionist to reduce human processing during actions, action observations and resulting cognitive tasks like action learning by observation to only mirror neuron function. Therefore, a broader term is needed.

1.2.4 The action observation execution network

The original definition of the AON by (Grafton, 2009) has only included action observation. To also stress the executive part of this network, Thomas et al. (2018) have used the term 'action observation execution network' (AOEN). In this thesis, I will refer to the human neural system being active when performing an action and when observing it as the 'action observation execution network' (AOEN). In this way, the execution of actions is also explicitly represented in the terminology used.

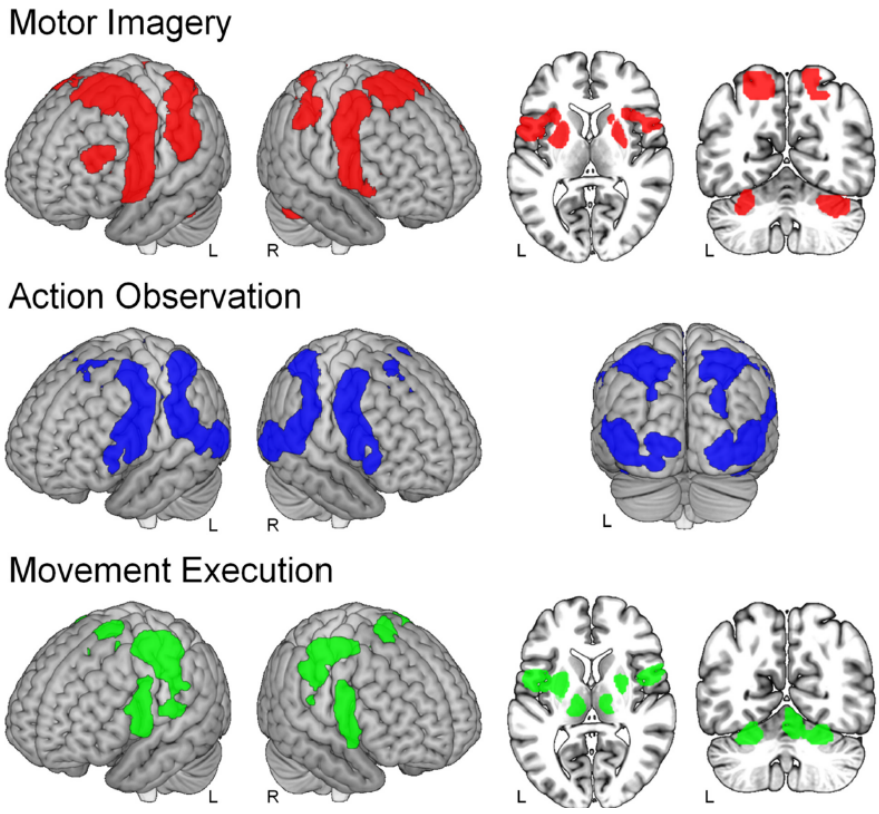


Figure 1.3: AOEN activations during motor imagery, action observation and movement execution. Results of the meta-analysis of Hardwick et al. (2018). Reprinted with permission from Hardwick et al. (2018), license number 4597621150368.

Figures 1.3 and 1.4 show the results of the most recent meta-analysis on action observation and movement execution (and motor imagery) performed by Hardwick et al. (2018) on human functional MRI data. For action observation, classically activated areas are - apart from the visual areas - the superior and inferior parietal lobe (SPL and IPL), the postcentral gyrus (primary somatosensory areas), the precentral gyrus (dorsal and ventral premotor cortex, vPMC and dPMC) as well as the inferior frontal gyrus (IFG), namely Areas 44 and 45 and parts of the posterior-medial frontal cortex (supplementary motor area (SMA) and pre-supplementary motor area (pre-SMA)). Apart from the

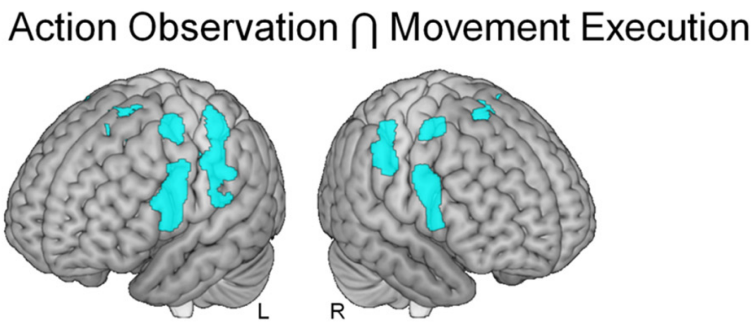


Figure 1.4: Conjunction of activations during action observation and movement execution. Results of the meta-analysis of Hardwick et al. (2018). Reprinted with permission from Hardwick et al. (2018), license number 4597621150368.

primary motor area (M1) classical movement execution areas are the dPMC and vPMC on the precentral gyrus, the IFG (esp. area 44), the postcentral gyrus (somatosensory areas), the posterior-medial frontal cortex (SMA) as well as the subcortical structures thalamus and putamen. Moreover, cerebellar lobules IV and V are reported.
For the conjunction between action observation and movement execution the following areas are reported: areas 3b and 2 of the postcentral gyrus, parts of the IPL, Area 44 on the IFG and the left and right precentral sulcus. Note that depending on atlas and/or examiner identification the extent of overlap of the vPMC with area 44 varies. Moreover, the right superior frontal gyrus, parts of the posterior-medial frontal cortex, as well as the right lobule VI of the cerebellum are reported.
If we want to consider the involvement of the whole cerebellum in these processes, it is important to note that traditionally, researchers have had to compromise between coverage, temporal and spatial resolution when utilizing fMRI. Therefore, often only the superior parts of the cerebellum were imaged. Additionally, older versions of SPM have used too small a bounding box, which excludes the inferior parts of the cerebellum (Abdelgabar et al., 2018).

1.2.5 Involvement and modulation of the action observation execution network in naturalistic contexts of experience

Overview of naturalistic learning and experience in the AOEN

Early fMRI studies in humans on the transfer between observation of actions and performance thereof have focused on the neural mechanisms of simple imitation. They used tasks like finger tapping (Iacoboni et al., 1999; Koski et al., 2003) to show common activations between observation, execution and combined observation and execution.
The first study on imitation-learning in humans using fMRI and a more naturalistic paradigm was performed by Buccino et al. (2004). As they pointed out, immediate imitation of simple actions that are already present in the motor repertoire like a finger tap cannot be considered as imitation learning. Consequently, they have deployed a more complex task: naive subjects had to observe the grip of different guitar chords on a guitar neck and reproduce this grip on a wooden guitar-like neck, both inside an MRI. This has been followed by a series of studies investigating similar tasks using MRI-compatible derivatives of musical instruments: subjects had to observe actions on these objects and reproduce them afterwards. Along these lines, researchers have looked into differences in brain activation along different contrasts: physical practice (Vogt et al., 2007; Higuchi et al., 2012; Sakreida et al., 2017), observational practice (Higuchi et al., 2012), musical expertise (Vogt et al., 2007; Sakreida et al., 2017), and reproduction of either rhythmical or sequential aspects (Sakreida et al. (2017)). Another non-musical observational learning study of manual transitive movements was published in 2006: Frey and Gerry (2006) investigated the observational learning of the assembling of a toy model inside the MRI. The subjects watched several videos of toys being assembled. After each video observation, participants were moved out of the MRI bore to assemble what they had just memorized, and then they were moved in again for the next video observation.
Another set of studies has investigated how experience and exposure shape activity in the AOEN when observing actions. Here, there has been no task of immediate imitation, meaning subjects do not need to memorize the moves shown in order to reproduce them. Typically, attention to the observation was ensured by either a judgment task ('How tiring is the movement?', Calvo-Merino et al., 2005) or 'How well could you reproduce the movement?' (e.g. Cross et al., 2006), or a visual matching task (Cross et al., 2012). In this line of research, the first studies focused on dance: Calvo-Merino et al. (2005) showed experience-related changes in the AOEN activity during observation of dance by employing different dance professionals (ballet and capoeira dancers) as well as naive subjects. This was followed by a study disentangling motor and visual familiarity by using sex-specific dance moves for professional ballet dancers of both sexes (Calvo-Merino et al., 2006). Further investigations have been changes due to deliberate

practice (Cross et al., 2006), observational practice by observing the movements (Kirsch and Cross, 2015) as well as observing someone else practice (Cross et al., 2017), and auditory practice (Kirsch and Cross, 2015).

Here is a (non-exhaustive) overview of studies utilizing naturalistic stimuli and actions to investigate AOEN activity during observation of actions, and which link this activity during observation to observational learning, experience, or exposure. All studies use either a classical fMRI block design or an event-related design. Using these naturalistic stimuli, the studies demonstrate the involvement of the AOEN in different aspects of experience and learning of actions and the observation of these actions. Most of the studies compare either action observation during different levels of professional experience or before and after a training period of several days. Immediate imitation is only examined by Buccino et al. (2004); Vogt et al. (2007); Frey and Gerry (2006); Sakreida et al. (2017). Of these studies, only Frey and Gerry (2006) examine the immediate reproduction of a longer sequence/chain of actions.

Whole-body movements

- **Dance studies:**
  - Calvo-Merino et al. (2005): Ballet, Capoeira dancers and naive subjects watched dance moves. Higher AOEN activation during observation of previously experienced/trained movements.
  - Calvo-Merino et al. (2006): Male and female ballet dancers observed sex-specific ballet moves by their own/another sex: Motor familiarity during observation due to their own previous practical experience leads to higher AOEN activation.
  - Cross et al. (2006): Ballet dancers observed and practiced new dance moves. Ability to execute a learned move is associated with increased AOEN activity during action simulation; rated own ability corresponds to IPL and PMC activity.
  - Cross et al. (2009): Naive subjects practiced a dance step game. Scanning before and after practice; compare physical, observational and no practice. Higher activity in parts of AOEN for observationally or physically practiced actions compared to unfamiliar ones after training.
  - Gardner et al. (2015): Dance stimuli shown to naive subjects, prediction of movements and familiarity ratings. More activity in IPL, IFG with higher familiarity, attenuated effective connectivity between parietal and temporal areas with increasing familiarity.
  - Kirsch and Cross (2015): Naive subjects trained dance with a Kinect video game - Compare observation of dance actions before and after different modalities of training (only observation, execution, auditory exposure). Higher AOEN response for more modalities, correlation between left PMC activity and participant's reproduction scores.

Manual movements

- **Guitar studies:**
  - Buccino et al. (2004): Imitation-learning study. Naive subjects observed and/or reproduced guitar chords inside the MRI machine. AOEN activity in action observation and imitation; BA46 activity during preparation of imitation.
  - Vogt et al. (2007): Experienced guitarists and naive subjects observed and/or reproduced guitar chords, some previously practiced, some not. Consistent AOEN activity; proposed model of DLPFC for action selection, combination and monitoring.
  - Higuchi et al. (2012): Naive subjects observationally or physically practiced guitar chords for four days, scanning during observation with imagery or with execution on day two and four. Connectivity of left prefrontal cortex and AOEN. Practice-related decreases of AOEN activity during observation with imagery. Observational and physical practice demonstrate similar AOEN activations during observation but higher practice-related decreases for physically practices chords.
  - Gardner et al. (2017): Naive subjects played guitar hero: comparison before and after training for observationally practiced (video) or manually practiced (guitar hero game) chords. Postulated cubic relationship for AOEN activity in relation to familiarity.
- **Toy model assembly:**
  - Frey and Gerry (2006): Naive subjects observed building of a toy model with the task of rebuilding it afterwards. Comparison between passive observation, attendance to exact sequence or outcome imitation. Higher AOEN activity for reproduction requirement than baseline and for exact sequence reproduction than for outcome reproduction. Performance related to IPL activity.
- **Knot tying:**
  - Tracy et al. (2003): Not a classical observational learning study as pictograph instructions were used, but a bi-manual naturalistic transitive learning task. Naive subjects needed to tie knots inside the MRI. Compared two scanning sessions with training in between. Posterior medial activity associated with good learning, weak learning with visuospatial activations.
  - Cross et al. (2012): Naive subjects learned to tie and/or name knots. Comparison of scans before and after training, visual matching task ('are these knots the same') during the scans. aIPS activity specifically for knots that participants had physically learned to tie.
  - Cross et al. (2017): Compared perceptual matching task of objects of two MRI sessions, before and after training: own manipulation, observing someone else manipulate. Interaction effect of training and scan session in part of SPL, right IPS and dPMC respond to trained knots.
- **Keyboard presses:**
  - Sakreida et al. (2017): Naive subjects and musicians observed, reproduced, or imagined either sequences or rhythms on a keyboard. Sequence reproduction leads to higher parietal and dPMC activity. DLPFC activity for both imitation tasks, postulated for cognitive control of imitation learning.
- **Surgery:**
  - Kok et al. (2018): Observation of surgical procedures by subjects of different skill levels. (In contrast to dance studies: only learned in adulthood, no developmental component. Eye-hand coordination task.) No significant results for the video x expert analysis.

Traditionally, there is a vast amount of literature on sequence learning in fMRI or, to a lesser extent due to spatial limitations in fMRI, on motor learning (i.e. force field learning). Most of these studies focus on changes of brain activity during the different stages of learning and production attempts. Consequently, a common way to test sequence learning in fMRI is to utilize simple finger presses on keyboards, the most common of these being the serial reaction time task (SRTT) developed by Nissen and Bullemer (1987). Based on abstract visual cues, subjects have to reproduce sequences as quickly and accurately as possible. For this sequence learning with several days of practice or training, activity increases as well as decreases also in motor processing areas have been reported as training effects (e.g. Steele and Penhune, 2010; Penhune and Steele, 2012; Wiestler and Diedrichsen, 2013). These could be associated with increased neural recruitment for trained behaviors as well as reduced activity due to higher neural efficiency (Wiestler and Diedrichsen, 2013).
As opposed to learning by following abstract cues, in this thesis the focus is on observing other people perform actions and learning to perform these by imitating the observed content.

AOEN involvement during short-term learning and the acquisition phase

Human skill learning can be characterized into an initial, fast learning stage, followed by a more sustained and longer-lasting, but slower learning stage (Dayan and Cohen, 2011). The initial stage of learning, the first acquisition phase, might seem like an easy subject to study, as one does not need to train the subjects or monitor them for a longer time. Surprisingly, there are currently few studies available that investigate these first trials of learning. Especially for procedural learning by observation, also referred to as imitation learning of more complex tasks, very few investigations have been performed. This is additionally surprising considering the vast amount of mirror neuron literature and also the fact that imitation learning or observational learning is a main part of human life (Bandura, 1971), with some researchers even hypothesizing that imitation learning is one of the main differences between humans and monkeys (Rizzolatti, 2014).
Of the studies listed in the previous section, only the following examine immediate reproduction after observation of a performance demonstration. The first study published in this line is Buccino's well cited investigation of imitating guitar chords (Buccino et al., 2004). Naive subjects had to observe an instructor showing guitar tabs on a fretboard and imitate the chord on an MRI-safe fretboard replica. In a follow-up, Vogt et al. (2007) investigated a similar setup with experienced and inexperienced subjects, who had also partly trained the chords before. In 2012, Higuchi et al. (2012) expanded this research by scanning on two days with practice in between and before, and also by investigating the connectivity between the dorsolateral prefrontal cortex (DLPFC) and the AOEN.

They could also affirm and establish that apart from the classical AOEN, the DLPFC is involved particularly in imitation learning (which has been suggested by Buccino et al. 2004). The last study in this line of musical reproduction is the one by Sakreida et al. (2017) who were using a similar task with a keyboard instead of a guitar fretboard, thereby approaching the literature on pure sequence learning tasks. On a non-musical approach, Frey and Gerry (2006) examined the observation and reproduction of the assembly of a toy model. They report a correlation between construction performance and activation during observation in the inferior parietal lobe (IPL). The other studies in this list focused on different aspects of novelty and observation, execution or imagery.
Taken together, there is some literature on the involvement and changes of AOEN activity in learning or experience related naturalistic contexts. Studies report a modulation of parts of the AOEN shaped by experience or by previous deliberate practice. Literature on the initial phase of learning remains sparse, but studies also report a fine-grained involvement here.

1.3 Motivation for this thesis

With this thesis, I would like to expand the knowledge about the processing of naturalistic stimuli and the involvement of the AOEN during this processing. Learning skills and sequences of actions by observation is pivotal for our everyday life - whether it be learning to ski, to knit, to play an instrument or to create a new cake. Modern media have established new social ways of observational learning by providing accessible videotaped instructions anywhere and anytime.
Previous experiments have often been highly controlled and therefore had a reductionist approach, which lowers the ecological validity. Naturalistic experiments ensure that our findings are transferable to everyday life situations - therefore I would like to present a naturalistic experiment on skill learning by observation. This requires new data-driven approaches, as the experiment will be less controlled. However, nowadays we have the computational power and knowledge - I will develop and adapt these methods for this thesis. The neural system mediating this process is thought to be the AOEN in the human brain - but other areas have also been discussed as being involved in learning or visuo-spatial processing. Consequently, to get an image of all involved processes, I intend to image the whole brain, to also include deeper cerebral and cerebellar areas reaching beyond the previously described AOEN areas. In this way, I will receive an exhaustive mapping of the cognitive processes involved. This requires me to use functional MRI as a go-to-method - and leaves me with the drawback of relatively bad temporal resolution in comparison to EEG or fNIRS. To improve this issue, I have made an effort to utilize the new fast multiband imaging in fMRI. Consequently, my aim for this thesis was to design an fMRI experiment that is as naturalistic as possible given the constraints of the machine. Given these, I will present a video-based approach analog to modern media.
The benefits of developing neurocognitive experiments and analysis methods for naturalistic skill learning are high: Of course, we gain knowledge about basic principles of learning. But also new information about cognitive processes acquired through a naturalistic video-based skill learning setup can inform any practical settings on e-learning and could later be adapted for diverse subject groups. Such a group could also be neuropsychiatric patients - especially since observational learning is a type of interaction and psychiatric illnesses can also be regarded as disorders of social interaction (Schilbach, 2016).

To design a naturalistic stimulus to probe skill learning, I thought of everyday situations where one can gain knowledge about procedures. With smartphones and available internet, one can search anytime for any kind of information. If I do not know how to change the tires of the car (as this is my first car), I will start to look for written information. However, probably I will quite soon search for a how-to video, which will

slowly show me the steps of the process. (And the research of Michas and Berry (2000), Wong et al. (2009) and Marcus et al. (2013) has shown that this is also a wise thing to do as it leads to more efficient learning.) I can watch this video to reproduce the steps, change the tires and save the garage bill - or I can decide to not do it (yet) because I have identified the actions and based on this decided that e.g. I am not able to do it - or that I should buy a torque wrench first. In all cases, I have watched the video and identified the actions - either to reproduce them or to perform different cognitive tasks based on the identification.

For the presented experiment, I have looked for a task which is more feasible to perform inside an MRI machine than changing tires. Handicrafts are a popular activity to learn and manual movements are doable inside a scanner. As the subjects are quite restricted in their position and bodily movements (especially shoulder movements) are very likely to cause head motion which severely corrupts the data quality, one aim was to find a manual task that is naturalistic but does not require much movement. Ideally, this means that the arms should stay fixed. Another aspect was that due to the position of the subjects, objects should not fall down as one cannot simply rise up to get them during the experiment. Lastly, metal objects are not possible inside the MRI. From this, I came up with a handicraft that is well-known, but is not practiced much in Germany: the folding of paper figures, also known as the folding of origami. Subjects are shown videos of paper foldings and should either memorize them to reproduce them or count the folds as a task of identifying actions without a need to learn. I believe that having an active control condition (like 'counting folds') is important: Previously, several control tasks simply involved watching a video/movie; but without a task, attentiveness cannot be guaranteed. Taken together, the second manuscript of this thesis explores the relation between the observers themselves during the action identification and action learning when watching origami videos.

Nevertheless, I also wanted to investigate the full process: in most cases there is a person performing these instructions. By watching these videos, effectively a modern version of communication between instructor and observer is established. Little literature exists on interaction with both parties being scanned (see chapter 1.1.3). Consequently, I wanted to investigate this relation as well. But I also had a further motivation from scientific literature: A characteristic of the AOEN is the activation during action execution and action observation, which has been previously examined on magnitude, not for temporal similarities. Therefore, I wanted to assess the commonalities between instructor and observer from this perspective. There is one problem though: if the instructor sees his/her own movement, one cannot rule out that a commonality is purely visual, meaning that the observed effect is merely the commonness of the own visual feedback and the (highly similar) observer's visual input. Consequently, I had to find a naturalistic task that the

instructor can do blindly - but is still able to do it in a way so that the observer does not notice anything unusual - because otherwise, the naturalistic context would be lost. I therefore searched for origami that can be folded blindly and then trained the instructor to do so. The first manuscript of this thesis examines the relation between the instructor and the observers.

Taken together, this thesis will investigate similarities between the instructor and the observers as well as between observers themselves when observing instructive paper folding videos - with either the task to learn and reproduce or the task to count folds.

2

Synchronization between instructor and observer when learning a complex bimanual skill

Kathrin Kostorz, Virginia L. Flanagin, Stefan Glasauer

Abstract While learning from an instructor by watching a 'how-to' video has become a common practice, we know surprisingly little about the relation between brain activities in instructor and observers. In this fMRI study we investigated the temporal synchronization between instructor and observers using intersubject correlation in a naturalistic setting of learning to fold origami. The blindfolded instructor was compared during action production to the observers during viewing of the instructor's video-taped actions. We demonstrate for the first time that the BOLD activity in the instructor's and observer's brain are synchronized while observing and learning a manual complex task with the goal of reproducing it, and rule out that this synchrony originates in visual feedback. Subjects who exhibited higher synchrony of brain activity with the instructor in the ventral premotor cortex while viewing the video for the first time were more successful in reproducing the origami afterwards. Furthermore, we show that changes in instructor-observer synchrony between the observational learning sessions reside in cerebellar areas, as well as differences in instructor-observer synchrony between learning and the non-learning control of counting folds. Since not only known motor production areas show synchrony, our results shed new light on the cerebellar involvement in action observation and learning.

2.1 Introduction

Learning manual skills, like painting, playing the guitar or performing handicrafts by observing another person is a natural human interaction. Learning about the world from other people plays an important role throughout the lifespan and is thought to be an efficient way of learning (Bandura, 1971; Frith and Frith, 2007). In today's society, natural human interaction has expanded far into the digital realm. Not only do we interact when video-chatting with friends, but unidirectional human interaction also occurs when we watch previously recorded online videos found on video platforms. Instructing other people this way has become commonplace: anyone can upload a self-made 'how to' video to the web. In turn, anybody interested can learn a skill by watching these online 'how to' videos. By watching the instructor in such a video, we gain insights into a specific process or skill. We gain information about spatial and temporal structures as well as motor aspects. Through watching, memorizing and imitating we can learn the process.
Learning actions by observing another person performing that action is thought to be mediated by a system referred to as the action observation network (AON) (Grafton, 2009) or the (overlapping) mirror neuron system (MNS) (Lago-Rodríguez et al., 2014). The term MNS originates in non-human primate research: the mirror neurons found in monkeys are active when the agent performs an action as well as when it sees that action being performed by another agent (see e.g. Rizzolatti et al. (2001)). In humans, typical action and action observation areas (not necessarily linked to findings of mirror neurons) include the premotor cortex (especially the dorsal and ventral parts), the somatosensory cortex, superior and inferior parietal areas, V5 and, if recorded, cerebellar areas (see Caspers et al., 2010 or Hardwick et al., 2018 for meta analyses). Thomas et al. (2018) introduce the term action observation execution network (AOEN) to human research, to emphasize both the execution and the observation of movements. For the purpose of this manuscript we will also refer to the network of brain regions active during actions and/or observation of actions as the AOEN, also to include brain structures that were not previously linked to the MNS. Moreover, for the purpose of this study, we will also use the terms 'imitation learning' and 'observational learning' interchangeably.
There is a vast literature on commonalities and differences of action observation and execution of single hand movements using functional magnetic resonance imaging (fMRI), see e.g. Molenberghs et al., 2012; Cook et al., 2014; Hardwick et al., 2018. Moreover, many studies investigate functional changes during the process of motor or sequence learning (see e.g. Hardwick et al., 2013). However, less studies address learning by observing another human's actions. Changes in activation in the AOEN during observational learning have been shown due to training and/or exposure. This was demonstrated in the context of dancing (e.g. Calvo-Merino et al., 2006; Kirsch and Cross, 2015), guitar playing (e.g. Buccino et al., 2004; Gardner et al., 2017) or knot tying (Cross et al.,

2012, 2017). Typically, these changes are shown by comparing different expert levels (e.g. Cross et al., 2006; Calvo-Merino et al., 2006) or different training levels in the order of magnitude of days to weeks (e.g. Cross et al., 2012; Gardner et al., 2017).
However, as Sakreida et al. (2017) have pointed out, the fMRI literature on the initial stage of imitation learning of skills/procedures is relatively sparse. In particular, learning through immediate imitation is rarely investigated. More often, during action observation subjects perform identification or rating tasks to judge difficulty or liking (Kirsch and Cross, 2015), or perform perceptual discrimination tasks (Cross et al., 2012). This might be related to the difficulty of performing movements inside the MRI scanner. To address this shortcoming, a series of studies have adressed imitation learning with immediate imitation in observing subjects; investigating guitar chord learning (Buccino et al., 2004; Vogt et al., 2007; Higuchi et al., 2012), keyboard sequences (Sakreida et al., 2017), or the construction of a toy model (Frey and Gerry, 2006). Nevertheless, except for the construction of a toy model, naturalistic bimanual imitation learning of object creation has not been investigated using fMRI, especially not in the initial learning stage. In particular, the neuro-cognitive relationship between observational learner and the instructing person remains unkown in this context. Given that performance and observation of actions are represented in the same brain areas in one person, is there a neural similarity between observer and instructor in these brain areas in a teaching setup? And how does this similarity relate to the learning performance of the observer? By answering our questions we are furthering our understanding of real world teaching scenarios.
There is a trend to investigate the neural processing of naturalistic stimuli to ensure more ecologically valid research (for a review see e.g. Hasson and Honey (2012)). In line with this, Hasson et al. (2012) stress the importance and need to investigate interactions in human neuroscience instead of only individuals' cognitive processes. Unidirectional naturalistic verbal communication with both transmitter and receiver being scanned has been investigated for story telling (Stephens et al., 2010; Silbert et al., 2014) and story re-telling (Zadbood et al., 2017). For interaction with manual actions, communication with gestures has been researched in a setting of a mimetic word guessing game (charades) (Schippers et al., 2009, 2010). Using a related method, functional near-infrared spectroscopy (fNIRS), manual action imitation and interaction of simple finger movements like finger-tapping (Holper et al., 2012; Dai et al., 2018) or object positioning (Bhat et al., 2017) have been studied, as well as an instructor-learner setting in the context of learning a song (Pan et al., 2018). Typically, during these interactive settings not only classical block-designs are employed and interpreted, but also similarities between subjects' brain activity, either on a voxel-to-voxel level (e.g. Stephens et al., 2010) or on a voxel pattern level (Zadbood et al., 2017). As these similarity findings show a close

relationship between the transmitter's and receiver's brain activity, Hasson and Frith (2016) argue for a coupling or a synchronization between the interacting parties instead of an alignment of unrelated neural activity.

However, the literature on human naturalistic interaction using fMRI with both participants' neural activity being recorded remains sparse, especially in a learning setup. To our best knowledge, the relationship between instructor's and observer's BOLD activity in the context of learning and reproducing a manual complex task has not been examined at all.

We investigated this relationship in a naturalistic manner by choosing the forming of an object which is popular and easily created inside an MRI machine: a paper figure, also known as origami. We have scanned an instructor performing the folding of an origami in an instructive way while being videotaped. This video was shown to subjects who observed it and memorized the foldings. The BOLD activity between the instructor performing and the observers watching was compared.

Given the findings of common activation between action execution and observation in classical fMRI block-design (Caspers et al., 2010; Hardwick et al., 2018), we would like to expanded these findings to temporal similarities. For pure action observation tasks without imitation or execution a temporal similarity or synchrony between observers has been shown (Nummenmaa et al., 2014; Thomas et al., 2018). Here, we hypothesize a similarity between instructor and observer brain activity in the AOEN over the course of the video. We postulate a synchronization of the time courses, corresponding to the fine grained changes of the video. To ensure that the common brain activation cannot originate from common visual processing, the instructor was blindfolded. Due to training, the instructor was able to perform the origami in the same way as having vision. This ensured that the naturalistic setting was kept. As Sakreida et al. (2017) have pointed out, observation with the task of immediate imitation might change the involvement of the AOEN. To compare the coupling between instructor and observer during action observation to imitate and during action observation to identify actions, we also produced a video of partly repetitive folds for which the subjects had to count the number of folds being made. For the main origami learning, we compare the instructor-observer similarity at the beginning and at the end of the learning process to see if the observers got more or less similar to the instructor.

To assess the similarity we utilize intersubject correlations (ISC) (Hasson et al., 2004). With the use of cross-correlations we investigated the temporal relation between instructor and observer brain activity. For different kinds of verbal and/or gestural communication, positive as well as negative lags were suggested (see Stephens et al., 2010, Zadbood et al., 2017 or Piazza et al., 2018). Given the autocorrelative nature of the signal we introduce a subsampling bootstrap routine (De Bin et al., 2016) to assess the lag.

Moreover, we were interested in the relation between instructor brain activity and receiver brain activity and subsequent observer performance. A relation between sender-receiver coupling and performance has been shown for listening comprehension (Stephens et al., 2010) or song performance (Pan et al., 2018). We hypothesized that a higher similarity of the observer to the instructor during action observation may be associated with better subsequent performance. Since the task is an immediate motoric reproduction, a better mapping onto own motor repertoire during observation may lead to better performance. The premotor cortex is thought to be involved in motor planning (e.g. Kantak et al., 2012), action selection and immediate goal encoding both in action production and action observation (Kilner, 2011). We therefore hypothesize that higher similarity of the observer to the instructor in the premotor cortex may lead to better performance. This would also be in accordance to mental simulation theory (e.g. Jacob and Jeannerod, 2005).

2.2 Materials and methods

2.2.1 Participants

One instructor and thirty-three observers were recruited for this study. The instructor was female, right-handed, 28 years old. She had no origami experience within the last 5 years, but practiced the required origami and the control folding sequence in the context of the experiment.

The 33 observers (15 female, mean age 27.2, range 21-36 years) were required to be right-handed, to have at least a high-school degree, and to be fluent in English as the experiment was performed in English. They did not have experience folding Japanese origami or any similar complex paper figures within the last 5 years. Rare folding of an already known easy traditional figure (simple plane, hat, boat, fortune teller) was tolerated. Moreover, participants were neurologically and mentally healthy, and did not have any further MRI contraindication. Handedness was assessed through 10 questions from the Edinburgh Inventory (Oldfield, 1971). All subjects had normal or corrected-to-normal vision. The participants were recruited through word-of-mouth, announcements at the university campus bulletin-boards, and at university-focused as well as expatriate-focused facebook groups.

All participants gave written, informed consent in accordance with the declaration of Helsinki and were reimbursed for their participation. The experiment was approved by the ethics committee of the medical faculty of the Ludwig-Maximilians-Universität Munich and the study followed all applying ethical regulations and guidelines.

After data collection, two subjects were excluded for reporting after data acquisition to have folded an origami sequence within the last 5 years. One person was excluded from the sample for reporting in the debriefing that she had moved her hands (counted with her fingers) during observation, even though observers were instructed not to do so. After inspection of the MRI data two more subjects were excluded from the sample; one because of abnormal brain anatomy, and one because of a head movement. The final sample that was analysed consisted of 28 subjects, of which 14 were female and mean age was 27.2 years (range 21 to 36 years).

2.2.2 Stimuli

We chose an origami butterfly with 8 - 9 folds (9 if an auxiliary crease was made during the last fold) as the stimulus for three reasons. First, the origami form is not well known in the Western hemisphere. Second, it can be folded blindly. Third, it is only possible to recognize the resulting butterfly figure after the last fold. This last reason avoided sudden comprehension in the middle of the folding. Participants had to remember the steps leading up the final origami form, they could not finish the origami by memorizing

the final shape at first observation.
To control for effects related to watching motor behavior without learning, we created a sequence of 8 unproductive folds (that do not result in any form or origami). The instructor made paper folds including the early folds of the butterfly origami. The starting fold was different to ensure that subjects did not suspect the same sequence again. The control folding sequence ended with an open sheet instead of a final origami. This way, the control was not predictable, had similar hand movements to the origami, but without learning to minimize the interference with the task of interest. Scrambled versions of the stimulus of interest have been frequently used as controls in naturalistic studies (Hasson et al., 2008b; Lerner et al., 2011; Regev et al., 2013; Silbert et al., 2014; Simony et al., 2016). Our control condition is similar, in that the folds of the original video we partially reproduced but in a different order and without a goal. The butterfly origami video shown in the main experiment had a length of 03:25 min, the video for the control condition of unproductive folding had a length of 03:32 min.

2.2.3 Experimental MRI setup

To be able to perform both the motor and visual components of the experiment inside the MRI scanner in the same location, we designed the following setup (see fig. 2.1. The key element was a standard Siemens MRI mirror system (MS), which has a double-mirror directed to the front/subject's feet and a standard mirror to the back of the bore. The conventional MS mount, which we used, is shiftable along the superior/inferior axis on rails attached to the head coil. Depending on the position of the MS, it offers a perspective either to the back or to the front of the bore. At the back of the bore, a 32" diagonal, 1920 x 1080 resolution LCD screen (NordicNeuroLab AS, 5009 Bergen, Norway) was installed, where the videos were displayed when the MS was shifted towards the back of the head to the "viewing position". The videos were digitally mirrored to be viewed correctly by the participant in the single mirror. The MS position to the front of the bore was the 'folding position'. We built a cardboard trapezoidal folding table that was placed around the subject's pelvis for folding. It had the right size and height to allow for people to observe their hands through front of the MS (no orientation change necessary due to the double mirror). The subject's arms were raised to the correct position for folding using foam pads.
An MRI-compatible infrared camera with VGA resolution (12M monochrome, MRC Systems GmbH, Heidelberg, Germany) was fixed to the mirror system pointing towards the folding table. This camera records in black-and-white and automatically adjusts the brightness, so we chose the origami paper to match the gray scale video and to minimize brightness variations throughout the video. Gray Japanese origami paper (tant, `http://www.kidstoyo.co.jp/` 15x15 cm, 80 g/m^2) was used for folding. The folds

could be seen easily at this size but the paper is still small enough to fold without big movements.
Potential head movements were of a large concern. Therefore, the experimental setup was designed and tested to minimize head movements. All participants were able to fold the paper keeping the wrists in the same position to minimize head movements through arm movements. Sandbags were placed on people's shoulders to additionally minimize arm movements. An inflatable head cap (Crania, Pearl Technology AG, Schlieren, Switzerland) was placed around the participants head and inflated to fill the gaps between a person's head and the headcoil, which dramatically reduced head movements. For switching between video observing and folding, subjects adjusted the MS themselves to their own personal optimal position for viewing and folding. The subjects were instructed to do so while keeping their head stable, and were reminded of this before every shifting. No MRI images were taken during movement of the MS.
Videos were recorded using the software Arrington ViewPoint PC-60 (version 2.9.5.117 Arrington Research, Inc., Scottsdale (AZ), USA) with a resolution of 320x240 px and a frame rate of 60Hz. Occasionally the frame grabber lost video frames. However, timestamps of the missing frames could be extracted from the accompanying logfiles. For missing timepoints, the previous or upcoming time-matched frame was filled into the resulting gaps using the VideoReader and VideoWriter functionalities of Matlab 2016b. This ensured an identical event timing and duration between the instructor and the observers. Afterwards, videos were digitally enhanced to show on the high resolution screen. Videos were scanner-synchronized and displayed to the observers using Matlab 2007b (The MathWorks, Inc., Natick (MA), USA) and Cogent 2000 (`http://www.vislab.ucl.ac.uk/cogent_2000.php`) on a personal computer.

2.2.4 Experimental procedure

Instructor sessions

The instructor was scanned in a separate MRI session before collecting the observers data. The instructor practiced folding both sequences inside and outside the scanner, to be able to fold in an instructive way. She deliberately showed the steps and moved her hands in a way that displayed the significant parts of the movements to the camera. She also folded at a moderate pace. To avoid visual sensory overlap, we required the instructor to perform both folding sequences without vision. Therefore, after practicing the folding with vision, the instructor practiced folding the sequences blindfolded. In the actual fMRI recording session, the instructor wore a sleeping mask to prevent visual feedback. She was instructed to leave her eyes open underneath the mask (it had bulges to allow for it) to maintain the highest similarity between instructor and observers. The instructor started folding in a self-paced way, after waiting for at least 20 scans (so that

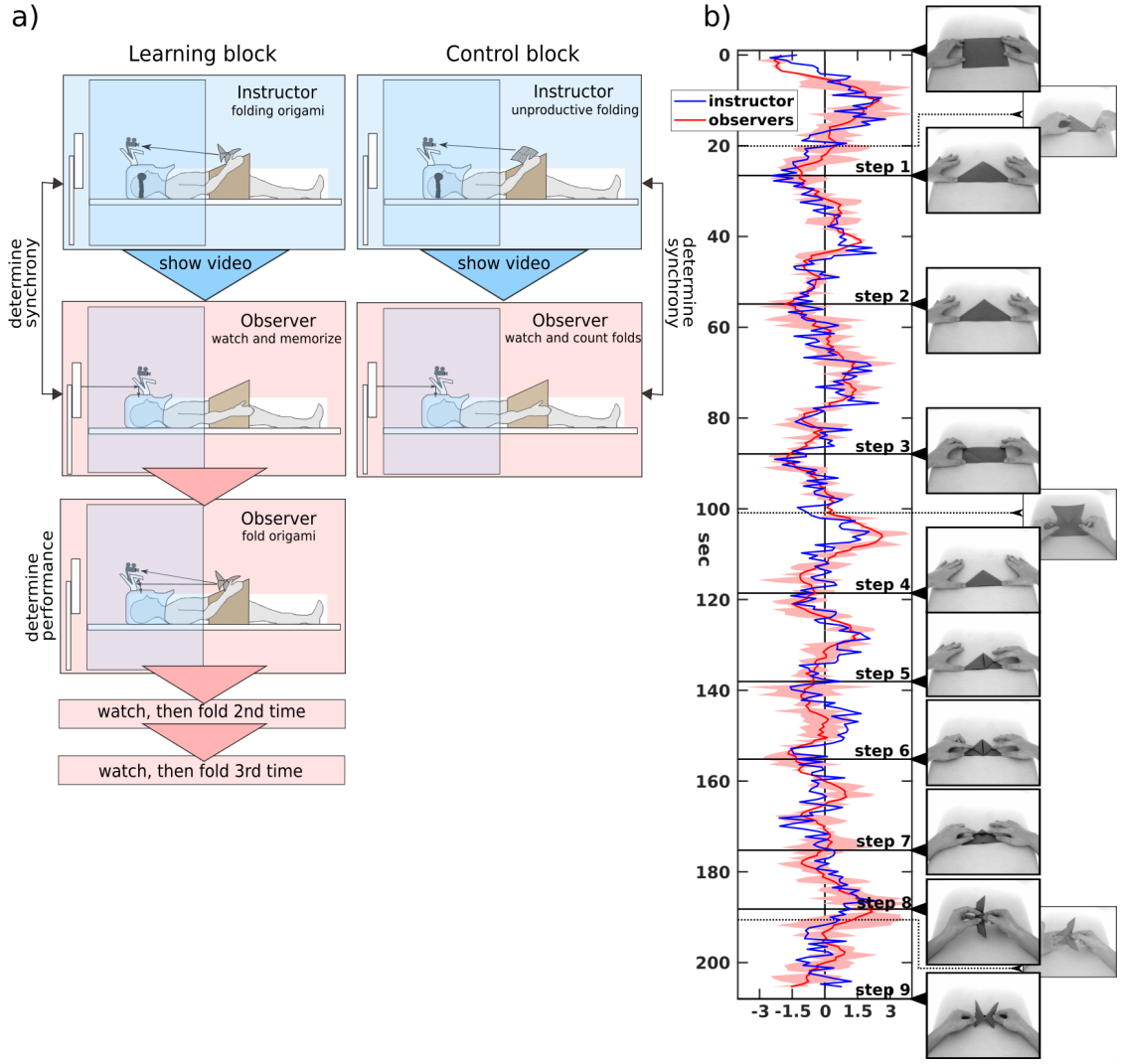


Figure 2.1: Experiment design - measuring brain activity during observational learning. a) The instructor video and fMRI data acquisition (first row, blue) was done in a separate MRI session. The best resulting videos were then shown to the observers (second and third row, red) with the task to watch and either memorize (learning condition) or count folds (control condition). After the learning condition, the observers attempted to fold the origami during fMRI data acquisition. The learning run was repeated two times to ensure that most people manage the origami. b) Instructor (blue) and observer (red) time courses overlaid for an exemplary voxel (-33,-42,63; average corr.=0.41) for the first learning run. (observers: solid line depicts the mean time course over all observers, shaded line the standard deviation). All time courses were standardized for display reasons. Solid black lines denote endpoints of folding steps. Images from the video were matched to the marked time point.

the first fold was after any T1 drifts in the MRI signal). We acquired two takes of the butterfly origami and chose the one that was better in terms of step visibility and speed. The same was done for the control sequence of unproductive folding.

Observer sessions

At the beginning of the main experiment, the observers were shown the final origami they were supposed to learn, by showing an image from the final stage of the video. Then the observers familiarized themselves with the setup, including manipulating the paper while viewing their hands through a mirror. In the final position in the scanner bore, they were instructed to fold anything they like with a piece of origami paper. Finally, subjects were told that, for the actual origami, they would not be judged on particular beauty but the folds must be well-recognizable. Our intent was to avoid subjects spending time remodeling their creases.

During the learning sessions, the observers viewed the instructive origami video three times (observation), and attempted to fold it once after each observation (reproduction). FMRI data were acquired during observation and reproduction, but only the observation sessions are analysed here. Before the start of each video, a black screen with a white fixation cross was shown to allow the T1 effects to saturate before the video started. No task was given during fixation. The task was to watch the video and memorize as far as you get. After each observation, the subjects took a new sheet of origami paper and reproduced the origami as far as they could get. Brain activity was measured and the hand movements were recorded with the same camera used for the instructors. As with the instructor, observers were asked to wait for 20 seconds after the fMRI data acquisition started before commencing to fold. Subjects gave a hand signal when they either thought they were finished or they did not know how to proceed any further. The subjects were told that they had an upper limit of 5 minutes to fold the origami. This limit was reached twice in the total 99 runs for all subjects. As the focus of this study was to investigate synchronization between production and observation of the same temporal sequence, only the observation runs of the observers will be assessed here.

During the control video of the unproductive folding, subjects were instructed to watch and count the number of folds made. Immediately after the fMRI run, they reported the number via the audio communication system of the MRI machine. We decided on this procedure instead of pure observation task for two reasons. First, it gives subjects the cognitive tasks of action identification and working memory that does not require memorization. Second, it ensures attentiveness and processing of the stimulus during the control video. The order of the control and learning runs were randomized. Subjects either started with 'watch and memorize' or 'watch and count folds', but were not interrupted during the three learning runs. The whole experimental work flow is depicted in

figure 2.1. Overall, scanning time added up to about an hour.
After the main experiment the experimenters did a debriefing with the subjects. This included open questions on how people memorized the steps ('please describe how you tried to memorize the steps while watching the videos') and a questionnaire with 7 unlikely items ('while you were watching the instructive videos, did you at any point think that:' e.g. 'the two hands belonged to two different actors' but also 'the person in the video could not see what he/she was doing', see supplemental information for full questionnaire, table 2.5) to assess if subjects found out about the blindfold. Of all participants only one person presumed that the instructor was blindfolded and one person was unsure. This way we can demonstrate that virtually no subject noticed the blindfold which might have led to unwanted cognitive processes during observation. Consequently, the naturalistic concept was kept.

2.2.5 fMRI acquisition and preprocessing

MRI data was acquired using a Siemens Skyra 3T scanner (Erlangen, Germany) with a 64 channel head coil. Before the fMRI runs, a field map (short echo time 4.9s ms, long echo time 7.38 ms) was recorded. Full brain coverage functional MRI data was provided using a multiband echo-planar imaging (EPI) sequence with a voxel size of 2.5x2.5x2.5 mm and a repetition time of 0.59 seconds (echo time 37.00 ms, flip angle 45 deg, 56 slices, no inter-slice gap, FoV 210 mm, matrix 84x84 px, phase encoding direction A/P, multiband acceleration factor 8). To allow for equilibrium effects with the short TR, the first 11 images were discarded from the analysis. During these scans, before the start of the video, a black screen with a white fixation cross was shown. For the folding sessions, the participant was told to wait for 20 seconds after the start of data acquisition before starting to fold, to ensure at least 11 images did not include folding. The butterfly origami run had 349 images and the unproductive folding run had 362 images. After functional imaging, an anatomical T1-weighted MPRAGE image was collected (TR=2.4 sec, TE=2.17 sec, flip angle 12 deg, slice thickness 0.75 mm, 256 slices, FoV 240 mm, matrix 320x320 px, phase encoding direction A/P, GRAPPA acceleration factor PE 2). Due to a technical error, the video of the control sequence started not at the beginning of an instructor scan, but with an offset of about 180 ms. We account for this during data analysis by linearly interpolating the instructor time course for this condition using MATLAB, align the instructor and observer scan times.

MRI data were preprocessed using the software SPM12 (7219) (`http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`) as well as FSL5.09 (`https://fsl.fmrib.ox.ac.uk/fsl`). Using SPM12, a voxel displacement map was created, followed by realignment and unwarping of the functional data, coregistration of the anatomical data to the

mean functional, and segmentation of the anatomical data. To ensure best alignment between the subjects, a common template of the instructor and observer anatomical data was created using DARTEL (Ashburner, 2007). Running the DARTEL algorithm and using this template, all subjects' data were normalized to MNI space with a voxel size of 3x3x3 mm, and a Gaussian smoothing kernel of 6x6x6 mm full-with at half maximum (FWHM). The smoothing kernel was kept small to ensure voxel-wise matching for the later ISC analysis, incorporating the recommendations of Pajula and Tohka (2014). FSL 5.09 was then used to highpass filter the functional data with a cutoff value of 100 ms. All functional images were inspected using the artifact detection tools (ART) version 2015-10 (`https://www.nitrc.org/projects/artifact_detect`). No motion artifacts were found using ART.
A gray matter mask from the subject cohort was created. It was computed by normalizing the segmented grey matter tissue probability map (TPM) of each subject, computing the median of all observers and thresholding at a probability of 0.4, as well as thresholding the normalized instructor TPM at 0.4. Additionally, signal loss was incorporated into this mask by computing the average signal intensity per brain voxel of the median subjects time course as well as the instructor time course and masking out voxels with 3 or more standard deviations from the average intensity. To improve computation time, only voxels within the gray matter mask were used in analysis.

2.2.6 Data analysis

For the comparisons between instructor and observer, three different types of analyses were used: 1) testing for overall synchrony, 2) testing differences in synchronization between runs, and 3) examining temporal shifts. Figure 2.2 shows a schematic of the procedure in each of these data analysis steps, and they are adescribed in detail below.

Temporal similarity between instructor and observers

To analyze the relationship between instructor and observer brain activity, we determined (cross-)correlations of the time courses for each corresponding voxel. This analysis is based on the intersubject correlation method (correlating at the same matching timepoint) which was first described for fMRI by Hasson et al. (2004), but using resampling methods. All similarity calculations were performed using custom MATLAB code (version R2018a). If not indicated otherwise, all calculations are performed at lag '0' meaning no time shift between instructor and observer activity.

Resampling methods for statistics The statistical significance of correlations and temporal lags were assessed through nonparametric resampling methods. Most standard statistical approaches assume independent observations, an assumption that is not up-

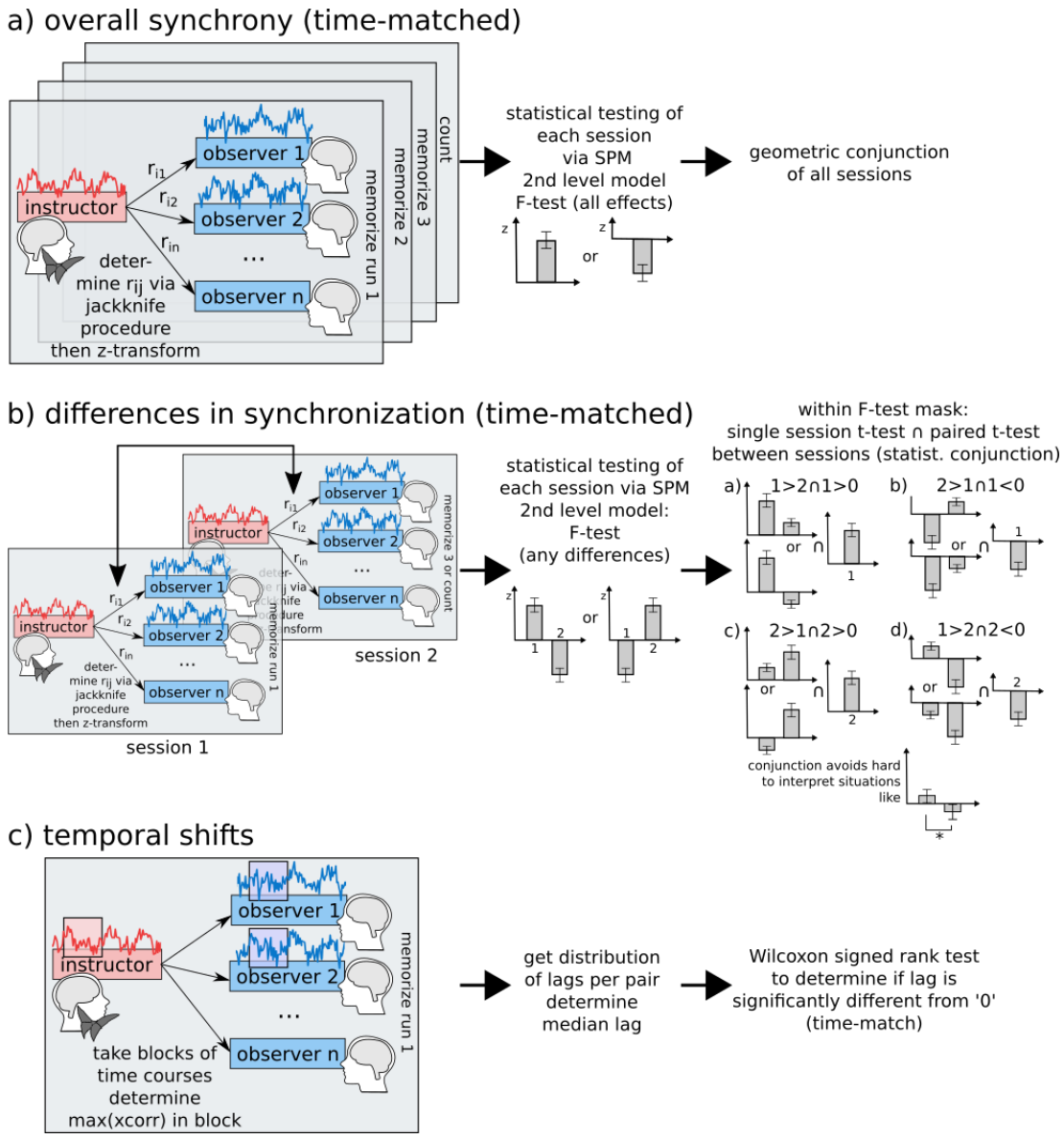


Figure 2.2: Data analysis workflow - the different instructor-observer comparisons. a)For the conjunction analysis (fig. 2.4) instructor and each observer were correlated. b)For the analysis of differences between sessions (fig. 2.5 and 2.6) the correlations calculated in a) were compared (see also materials and methods). c)The temporal shifts (fig. 2.8) were assessed via a block bootstrap routine (see 2.2.6).

held in a fMRI time series. Especially with the high temporal resolution used here, the autocorrelative structure of the data violates this assumption. (See figure 2.9 for illustration.) To resolve this problem, alternative methods based on jackknife or bootstrap techniques have been proposed. For the present problem we used two closely related methods, the jackknife (Abdi and Williams, 2010a) method for the correlation coefficients and the subsampling method (Berg et al., 2010) to assess the temporal lags. The jackknife method circumvents some of the issues mentioned above, since it uses the original data in their given order, leaving out a single pair of values per iteration. The subsampling method estimates the lags based on sampling fixed length subsets of the original values (block sampling or m-out-of-n bootstrap). Within each subset block the maximum of the cross-correlation is then computed. This results in a distribution of lags with as many lags as subsets/blocks. The median of this lag distribution is then used for further analysis. One of the issues in subsampling is the correct choice of block length (De Bin et al., 2016). We chose the value proportion 1/e (approx. 37%) given by Peterson et al. (1998) as our block length. This block length gave good results for our tests using synthetic i.i.d. data as a control. To compare our temporal delays to the classical procedure of taking the peak of the cross-correlation (e.g. Kreuz et al., 2007), we plotted the median delays over all subjects and voxels for both methods in figure 2.12.

Similarity To assess similarity (in a and b of figure 2.2) we computed the voxelwise Pearson correlation coefficient for each subject pair without a timeshift. The correlations were calculated via the jackknife routine and then z-transformed via the Fisher transformation. This resulted into an observer-instructor correlation image for each of the 28 subjects and for each condition.

Temporal lags Using the subsampling routine, we determined an optimal lag between each subject and the instructor for each voxel. In our case we investigated temporal shifts between -15·TR and +15·TR (-8.85 sec to +8.85 sec). To assess whether these lags were consistently different from lag '0' in one direction for all subjects (i.e. on a group level the observers' BOLD activities either lag behind instructor's activity or precede it), we tested the lags per voxel using the two-sided Wilcoxon signed-rank test under the null hypothesis that the median is '0'. To test for significant differences between lags between different sessions, we performed this as a two-sided paired test. Note that in previous studies reporting potential temporal shifts, the lag was determined using the maximum of the average cross-correlation (Zadbood et al., 2017; Piazza et al., 2018). However, to be sure that the lag exists, it is important to statistically test whether this lag is significant. This means statistically testing if the lag is different from '0', or whether there is an overall positive or negative lag in the lag distribution. The statistical regression-contrast approach of Stephens et al. (2010) could not be adapted for our data as we

were sampling higher and consequently have increased autocorrelations. In the work of Stephens et al. (2010), only the independence of the directly neighboring regressors was violated. With our higher sampling rate, we could not ignore this assumption as more neighboring timepoints were correlated (compare fig. 2.9 to the corresponding figure in the supplementary information of Stephens et al. (2010)). Therefore, we could not test each shifts's correlation separately or perform a contrast analysis on time points as independent regressors. Instead, we bootstraped a lag distribution based on subsamples per time course and then tested for statistical significance.

Statistical parametric mapping

The statistical significance of synchronization was tested using SPM12. The Fisher z-transformed correlation images were loaded into a group level model. To assess significance of each coupling we used a t-test design. For the comparison between learning and counting we used a paired t-test design, for the comparison between the learning runs a repeated-measures ANOVA design. Note that for this data set, both positive and negative correlations between the observers and the instructor are possible. Therefore, we resorted to the following procedure: Overall significance for all conditions was first tested using an F-contrast which is essentially a two-sided t-test. This way, any effect of difference (from '0' or from the other condition in case of the paired test) was found, regardless of the direction. (For statistical thresholding see next paragraph.)
For the conjunction of all conditions we then created a conjunction mask of each thresholded image using the fslmaths function of FSL to optain only voxels significantly coupled between instructor and observers in all conditions. For illustrative purposes we display the mean z-value of each voxel averaged over all conditions and subjects.
However, the F-contrast does not provide any information about the directionality of the effect. Therefore, when comparing between sessions we computed additional t-contrasts: First, we created a mask of the significant F-test; within this mask every further test was performed. This way, all voxels which are significant at the end of this procedure show significant difference between the sessions based on the F-test. In principle, we can have positive or negative correlations for both conditions. To determine whether condition one has a stronger positive effect than condition two, it is important not only to test for run '1>2' but also make sure that '1>0'. This way we can avoid cases in which'1<0' but run 1 is less negative than 2, where we would rather attribute a stronger coupling to run 2 than run 1. This leaves four tests: '$1 > 2 \cap 1 > 0$', '$2 > 1 \cap 2 > 0$', '$1 < 2 \cap 1 < 0$' and '$2 < 1 \cap 2 < 0$'. Note: with this categorization it is possible for the same voxel to be significant in two conditions if the two conditions have a comparable effect size that is different from '0' in opposite directions. However, we do not find any overlapping effects, which means we can attribute our results to a stronger coupling in one condition

compared to the other.

Statistical thresholding

For the statistical thresholding of the SPM analyses, we followed the ideas and procedure described in Thomas et al. (2018). In brief, we took the maximum statistical threshold of $p_{uncorr.}<0.001$ and $q_{FDR}<0.05$, and then set a cluster correction threshold of $p_{cluster,FWE}<0.05$ determined at $p_{uncorr.}<0.001$. This way voxelwise significance was at the level of $q_{FDR}<0.05$ and clusterwise significance at $p_{FWE}<0.05$, thereby avoiding loose FDR corrected thresholds. When testing for significant instructor-observer-coupling against '0', we found $p_{uncorr.}<0.001$ to be stricter than $q_{FDR}<0.05$ and $q_{FDR}<0.01$ in all conditions.
We thresholded all SPM statistics at a cluster extent of 24 voxels (648 mm^3). This was determined by identifying each statistic's cluster correction extent threshold (determined at $p_{uncorr.}<0.001$ with SPM12) and then taking the maximum of all extents. This way every statistics performed by SPM presented in this paper survives a cluster threshold of $p_{cluster,FWE}<0.05$. For the non-parametric test of the lags (Wilcoxon signed-ranks), we plot $p_{uncorr.}<0.001$ values threshold at $k>24$.Between-session contrasts (both F-contrasts and t-contrasts) were thresholded at the statistical level described above.The anatomical locations reported in this paper were determined using the SPM Anatomy Toolbox version 2.2b. The highest submaxima were reported here.

2.2.7 Behavioral rating of the observer performance

Behavioral scoring was performed by two examiners during the data collection process, prior to data analysis. The performance of the observers was judged by adding up points for correct steps. Nine steps were counted in total to allow a more detailed rating based on the auxillary crease some subjects made in the last fold. One point was given for each correct step respectively fold. Zero points wwere given for a missing step. A half a point was given for each step that was clearly folded but incorporated a flaw. This could be a fold that was made in the correct shape, but in the wrong place, or a 3D fold that was folded correctly, but did not 'stick in place'. In total, this resulted in possible scores between 0 to 9 points. The behavioral rating changes across runs are used to motivate the comparison of the instructor-observer-synchrony between the first and the last learning run. Moreover, the ratings are correlated with the instructor-observer-synchrony in the defined ROIs (see section below) and also exploratorily voxel-wise for the whole brain.

2.2.8 ROI analysis

We hypothesized that subjects who perform better in the immediate reproduction may correlate better with the instructor in the corresponding premotor areas. To construct the region of interests (ROIs) we took the premotor peak coordinates of the conjunction analysis of 'action observation' and 'action production' in the meta-analysis of Hardwick et al. (2018). Using MarsBaR version 0.44 (`http://marsbar.sourceforge.net/`) we extracted spherical ROIs around these coordinates with a radius of 9mm. At this radius the left dorsal premotor cortex ROI also includes the (only) peak of the meta-analysis for sensorimotor and sequence learning reported in Hardwick et al. (2013). Since origami folding involves sequence learning and sensorimotor aspects we hypothesized this specific ROI to show behavioral performance correlates with synchrony to the instructor.
As control ROIs, we chose the primary auditory cortex as we do not expect a correlation with performance here. To also receive comparable spherical ROIs we extracted the peak voxels on both hemispheres of the cytoarchitectonic map 'Auditory_Te10' of the SPM anatomy toolbox (version 2.2b). We then created spherical ROIs around these peaks the same way as the premotor ROIs. All of the spherical ROIs were conjuncted with the above described grey matter mask to censor the parts of the spheres which are outside of the voxels that were analyzed. Within the resulting areas we averaged each subject's time course and the instructor time course. We then calculated a Pearson correlation between each subject's ROI time course and the instructor's ROI time course. To determine the relation between the instructor-observer coupling/correlation and the behavioral scores a Spearman rank correlation between these two values was computed.

2.3 Results

2.3.1 Behavioral results

As a measure of performance and learning, participants were asked to fold the same origami figure that they watched being folded immediately thereafter and their behavior was recorded with a video camera and rated by two independent raters. As seen in Figure 2.3, the origami has nine folds in total. Performance was rated at each fold based upon whether the participant made the correct fold (1 step), did not make the correct the fold (0 steps) or made a partially correct fold (0.5 steps). A sum across all the folds was then the measure of performance with a maximal value of nine steps for completely finishing the origami.

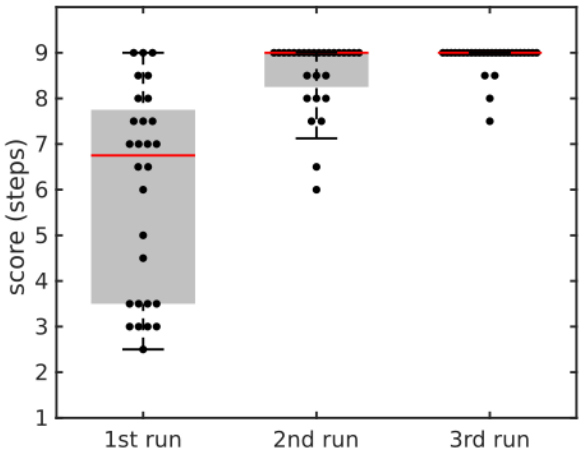


Figure 2.3: Folding performance after each observation run, depicting the learning process. While for the first run, performance was variable between the subjects, they learned to fold the origami almost completely over the course of the experiment. The red bars denote the median performance and black asterisks correspond to individual subject performance. The whiskers are defined as lower/upper quartile $\pm 1.5\times$ the interquartile range.

All of the 28 subjects analyzed were able to perform more than seven of the nine folds to create the figure, and all except for 4 subjects managed to fold the entire origami within the three runs (see fig. 2.3). Folding performance after the first run was highly variable: two people completed the origami already after the first view but nine failed to complete the first 3D fold (step 4). After viewing the video a second time, all subjects folded past the first 3D fold, six people managed at least the first part (step 8) of the second 3D fold (step 8 and 9) and 18 subjects completed the origami. This means that 24 out of 28 subjects finished or almost finished the origami after the second attempt.

For this reason, we chose to focus on the difference between the first and the third run to examine the effects of learning.
In the control condition, in which participants counted the number of fold, but no origami figure was made, performance was based on whether the participants correctly counted the number of folds that were made. All except one of the subjects counted the number of folds correctly.

2.3.2 Synchronization between observers and instructor

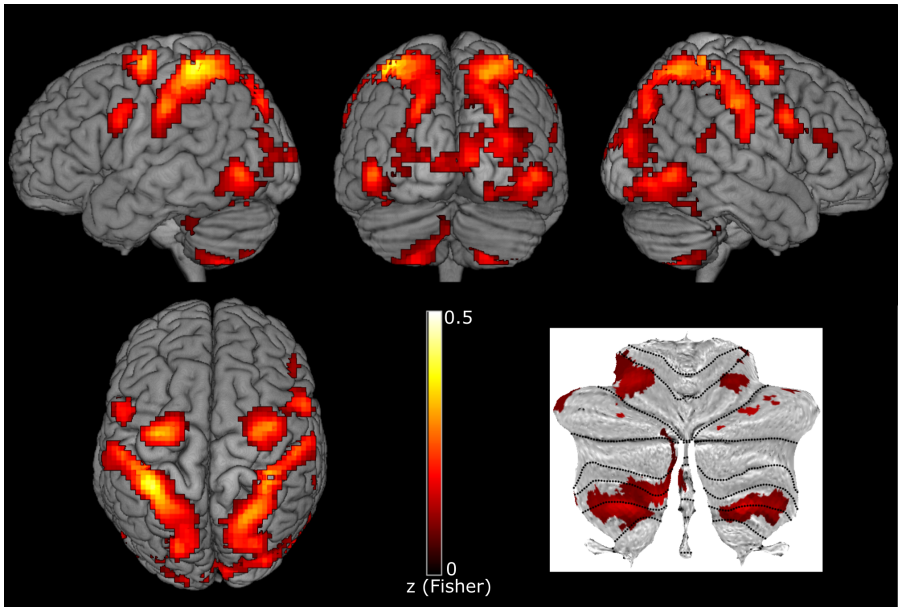


Figure 2.4: Synchrony across all conditions. The AOEN is consistently synchronized between instructor and observers in the three 'memorize' conditions and the 'count folds' condition, tested via a conjunction. This includes the cerebellar hand motor areas. Each condition is thresholded at $p_{uncorr}<0.001$ and $k\geq24$ before the conjunction. (Note that $q_{FDR}<0.01$ is less strict). For illustration purposes, the average correlation (Fisher-z transformed) over all conditions for each voxel is presented and masked with the conjunction mask of all four significant conditions. Data are overlaid onto mricron's ch2better template and the suit toolbox's cerebellar flatmap respectively.

To assess the level of synchronization in individual brain areas across all of the conditions tested, we correlated the time courses of corresponding voxels between the observers and the instructor at lag '0' (see materials and methods). Figure 2.4 and table 2.1 show a conjunction of significant instructor-observer couplings of all four experimental observation runs (three origami observations and one counting). The AEON showed significant correlations across all conditions, including the premotor cortex, somatosensory cortex, inferior and superior parietal cortex; and cerebellar locations, particularly lobules VI and VIIIa and VIIIb. M1, another classical motion execution area, did not show similarity between observers and instructor at our statistical threshold. Interestingly, significantly correlated voxels were also found in the visual cortex for all conditions, although the instructor was blindfolded.

Individual runs also showed extended synchronization between instructor and observers: significant correlations were found in parts of the supplementary motor area (SMA) or pre-SMA (first, second memorize and count run), both parts of the AEON. Clusters with significant correlations were also found in anterior and mid cingulate areas

(first memorize run). Subthreshold, the left primary motor areas show small anticorrelations. All significantly correlated voxels for each of the four experimental conditions are presented in Figure 2.10.

location	cluster size	peak statistics x	y	z	stat.	peak location *area with max probability*
conjunction of all conditions						
R Postcentral Gyrus, R Superior Parietal Lobule, R Inferior Lobe, R Occipital Cortex	1255	33	-36	57	0.36	R Postcentral Gyrus; Area 3b 48%
		27	-57	63	0.31	R Superior Parietal Lobule; Area 7A (SPL) 37%
		21	-60	60	0.31	R Superior Parietal Lobule; Area 7A (SPL) 64%
		48	-21	39	0.31	R Postcentral Gyrus; Area 3b 43%
		33	-42	66	0.31	R Postcentral Gyrus; Area 1 47%
L Postcentral Gyrus, L Superior Parietal Lobule, L Inferior Parietal Lobule	1070	-36	-42	63	0.42	L Postcentral Gyrus; Area 2 40%
		-36	-33	45	0.32	L Postcentral Gyrus; Area 2 20%, Area 3a 17%
		-54	-21	39	0.29	L Inferior Parietal Lobule; Area 1 42%
		-48	-30	54	0.28	L Postcentral Gyrus; Area 2 38%
R Occipital Cortex	327	48	-60	-3	0.25	R Middle Temporal Gyrus; Area hOc5 [V5/MT] 20%
		45	-75	-9	0.22	R Inferior Occipital Gyrus; Area hOc4la 64%
		24	-87	-15	0.15	R Lingual Gyrus; Area hOc3v [V3v] 58%
		36	-81	-12	0.15	R Inferior Occipital Gyrus; Area hOc4v [V4(v)] 64%
L Occipital Cortex, L Cerebellum (VI)	313	-48	-69	-3	0.26	L Inferior Occipital Gyrus, Area hOc4la 49%
		-27	-51	-24	0.16	L Cerebellum (VI); Lobule VI (Hem) 94%
L Precentral Gyrus, L Superior Frontal Gyrus	237	-33	-9	66	0.34	L Precentral Gyrus
R Precentral Gyrus, R Superior Frontal Gyrus	226	36	-6	63	0.29	R Superior Frontal Gyrus
L Cerebellum VIIIa and VIIIb	216	-21	-57	-57	0.19	L Cerebellum (VIII); Lobule VIIIb (Hem) 70%
		-9	-72	-48	0.16	L Cerebellum (VIII); Lobule VIIIa (Hem) 78%
R Precentral Gyrus (inferior)	150	57	9	33	0.25	R Precentral Gyrus; Area 44 29%
R Cerebellum VIIIa and VIIIb	112	21	-57	-54	0.18	R Cerebellum (VIII); Lobule VIIIb (Hem) 77%
		30	-45	-51	0.12	R Cerebellum (VIII); Lobule VIIIa (Hem) 59%
L Precentral Gyrus (inferior)	109	-51	6	30	0.22	L Precentral Gyrus; Area 44 39%
R Temporal Parietal Junction	32	60	-45	12	0.14	R Superior Temporal Gyrus
		60	-39	24	0.12	R SupraMarginal Gyrus; Area PF(IPL) 38%
R Inferior Frontal Gyrus	25	51	33	18	0.12	R IFG (p. Triangularis); Area 45 38%
R Cerebellum VI	22	24	-51	-24	0.11	R Cerebellum (VI); Lobule VI (Hem) 86%

Table 2.1: Table of locations for the synchrony across all conditions (each voxel has a significant instructor-observer correlation for each run). This table corresponds to figure 2.4. For statistical thresholding please refer to materials and methods. For illustration purposes, the statistics ('stat.') reported here are Fisher z-values averaged over all subjects and conditions within the mask. x, y and z demark MNI coordinates in mm. (The size of next smaller cluster is 11 voxels).

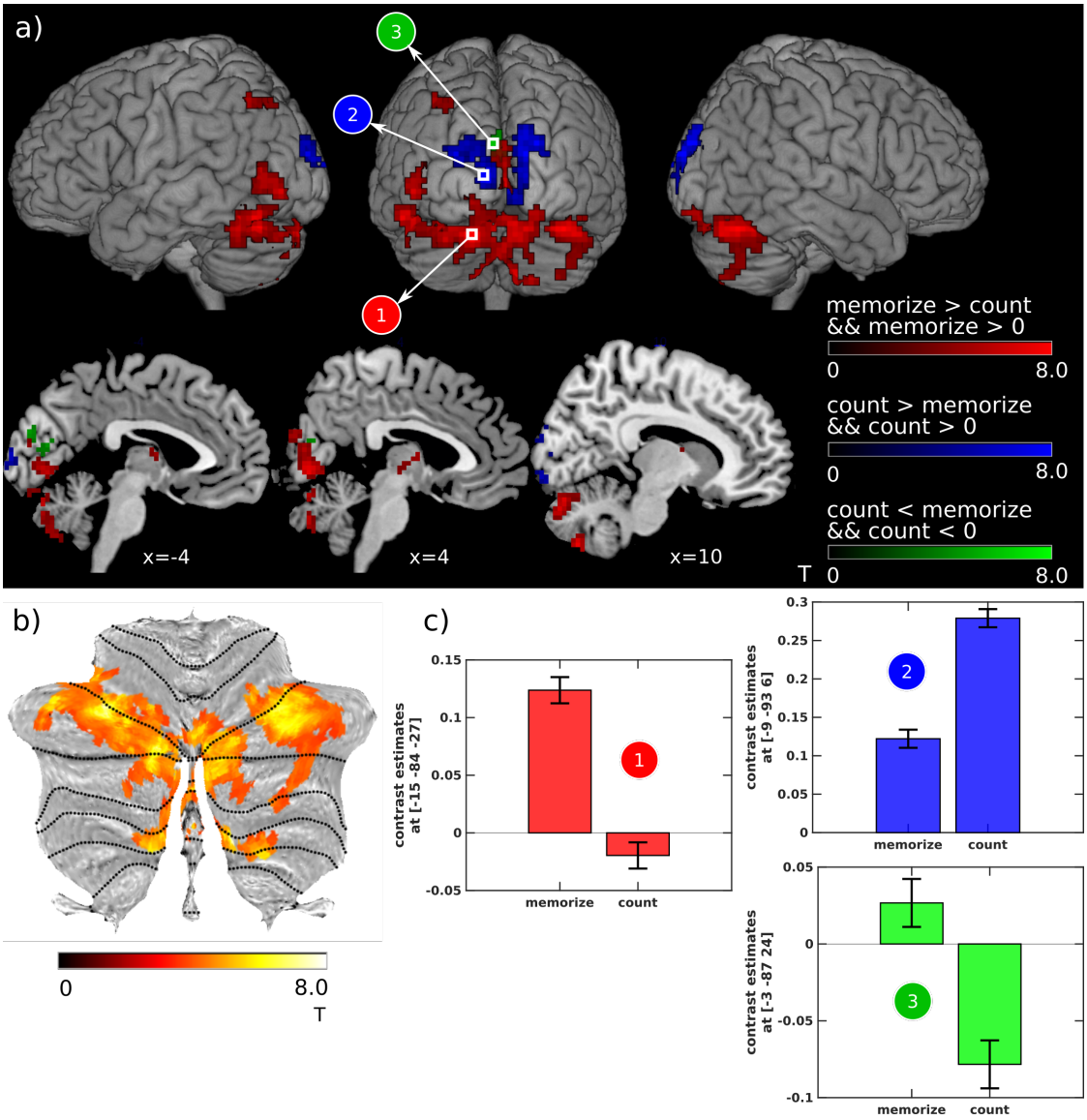


Figure 2.5: Differences in instructor-observer correlation between conditions. Higher instructor-observer-synchrony is seen in cerebellar areas during the learning run compared to the count folds run. See section 2.2.6 for the thresholding procedure. The t-statistic is displayed with T_{thr}=3.42. a) significant differences between runs separated by directionality (see materials and methods). b) cerebellar flatmap depicting the higher correlation for memorization than for counting, equivalent to the cerebellar part of the red heatmap in a). c) contrast estimates for exemplary voxels for each condition (color coding analog to a)). Error bars show standard errors of the mean. Data are overlaid onto mricron's ch2better template and the suit toolbox's cerebellar flatmap respectively.

location	cluster size	peak statistics x	y	z	T	peak location *area with max probability*
(first watch to memorize > count folds) ∩ (first watch to memorize > 0) (positive correlations)						
R+L Cerebellum	664	-15	-84	-27	8.92	L Cerebellum (Crus 1); Lobule VIIa crus I (Hem) 93%
		-33	-66	-24	8.90	L Cerebellum (VI); Lobule VIIa crusI (Hem) 54%
		-54	-63	-18	7.41	L Inferior Occipital Gyrus; Area FG2 51%
		9	-69	-51	7.40	N/A; R Lobule VIIIa (Hem) 40%
		15	-78	-24	6.90	R Cerebellum (Crus 1); Lobule VIIa crus I (Hem) 51%
		9	-78	-21	6.87	R Cerebellum (VI); Lobule VI (Hem) 95%
		9	-81	-27	6.76	R Cerebellum (Crus 1); Lobule VIIa crus I (Hem) 79%
		-48	-54	-18	6.71	L Inferior Temporal Gyrus; Area FG4 45%
		-30	-75	-30	6.37	L Cerebellum (Crus 1); Lobule VIIa crusI (Hem) 91%
R Cerebellum	237	36	-66	-24	8.05	R Cerebellum (VI); Lobule VI (Hem) 74%
		42	-75	-27	7.03	R Cerebellum (Crus 1); Lobule VIIa crus I (Hem) 96%
		42	-57	-24	7.00	R Cerebellum (VI); Area FG2 45%
		27	-63	-21	5.49	R Cerebellum (VI); Lobule VI (Hem) 92%
		33	-75	-51	4.57	R Cerebellum (VII); Lobule VIIa crus II (Hem) 80%
		48	-57	-30	4.39	R Cerebellum (Crus1); Lobule VIIa crusI (Hem) 72%
		39	-66	-45	4.27	R Cerebellum (Crus 2); Lobule VIIa crus I (Hem) 72%
		42	-66	-36	4.24	R Cerebellum (Crus1); Lobule VIIa crus I (Hem) 89%
R+L Occipital Cortex	118	3	-75	-3	7.32	R Lingual Gyrus; Area HOc1 [V1] 60%
		0	-78	0	7.15	L Lingual Gyrus; Area hOc1 [V1] 53%
		3	-78	12	5.59	L Calcarine Gyrus; Area hOc2 [V2] 51%
		3	-84	21	5.12	L Cuneus; Area HOc2 [V2] 31%
L Occipital Cortex	76	-48	-78	-3	6.75	L Inferior Occipital Gyrus; Area hOc4la 86%
		-39	-63	9	6.22	L Middle Occipital Gyrus
		-42	-66	3	5.76	L Middle Occipital Gyrus
		-48	-75	3	5.46	L Middle Occipital Gyrus; Area hOc4la 66%
Thalamus	25	6	-3	9	5.55	N/A; R Thalamus: Temporal 58%
		6	-3	3	5.06	N/A; R Thalamus: Temporal 68%
		-3	-3	9	4.89	N/A
		3	-15	-3	4.17	N/A; Thalamus; Temporal 43%
		0	-12	3	4.13	N/A; Thalamus; Temporal 26%
L Inferior Parietal Lobule	24	-36	-60	48	5.38	L Inferior Parietal Lobule; Area hIP3 (IPS) 40%
		-30	-69	45	4.31	L Inferior Parietal Lobule
(count folds > first watch to memorize) ∩ (count folds > 0) (positive correlations)						
R Occipital Cortex	115	21	-90	24	9.04	R Superior Occipital Gyrus; Area hOc4d [V3A] 41%
		15	-96	18	8.27	R Superior Occipital Gyrus; Area hOc2 [V2] 52%
		15	-90	30	6.66	R Cuneus; Area hOc4d [V3A] 62%
		15	-87	6	6.27	R Calcarine Gyrus; Area hOc1 [V1] 78%
		15	-96	9	5.88	R Cuneus; hOc2 [V2] 38%
		6	-93	-3	5.85	L Calcarine Gyrus; Area hOc1 [V1] 74%
		15	-87	18	5.82	R Cuneus
		15	-96	3	5.67	R Calcarine Gyrus; Area hOc1 [V1] 66%
L Occipital Cortex	73	-9	-93	6	9.38	L Superior Occipital Gyrus; Area hOc1 [V1] 45%
		-6	-99	3	8.31	L Calcarine Gyrus; Area hOc1 [V1] 58%
		-24	-93	18	7.28	L Middle Occipital Gyrus; Area hOc4d [V3A] 39%
		-21	-96	15	5.91	L Middle Occipital Gyrus; Area hOc3d [V3d] 46%
		-15	-93	15	5.89	L Superior Occipital Gyrus; Area hOc4d [V3A] 50%
		-9	-102	12	5.82	L Superior Occipital Gyrus; Area hOc2 [V2] 52%
(count folds < first watch to memorize) ∩ (count folds < 0) (negative correlations)						
L Occipital Cortex	31	-3	-87	24	4.76	L Cuneus; Area hOc3d [V3d] 48%
		-3	-75	15	4.57	L Calcarine Gyrus; Area hOc2 [V2] 42%
		0	-72	18	4.47	L Calcarine Gyrus; Area hOc2 [V2] 28%
		-6	-81	9	3.96	L Calcarine Gyrus; Area hOc1 [V1] 57%
(first watch to memorize < count folds) ∩ (first watch to memorize < 0) (negative correlations)						
n.s.						

Table 2.2: Clusters found in the various contrasts between the first origami run and the counting condition. Table corresponds to figure 2.5. Statistical procedure described in materials and methods. In the last condition no significant voxel was found.

2.3.3 Instructor-observer correlations during observational learning

To assess the specific effects of learning compared to counting, we compared the first memorization run with the control run. These two runs were counterbalanced and both stimuli were novel at the time of display. The biggest differences between these runs were seen in cerebellar and visual areas. A distinctive testing, based on the logical "and", was necessary to test for specific effects (see materials and methods). For example – to test for higher synchrony in memorizing the origami compared to counting (red in Figure 2.5), it was necessary to test for memorize > counting AND memorize > 0, so that we tested for higher correlations during the first memorize run compared to counting only when the correlations during the first memorize run were positive. Figure 2.5 and table 2.2 show the resulting effects. Higher positive correlations between subjects and instructor during the first memorization run were seen in visual areas and the cerebellum, as well as in a small cluster in the left superior parietal lobule. The biggest clusters were located in cerebellar lobules crus 1 and crus 2, stretching into lobule VI. Smaller clusters were located more inferior in cerebellar lobules VIIIa and b. The clusters in the visual areas incorporated primary visual areas and the left V5/MT. Higher positive correlations between instructor and observers during the 'count folds' condition was found in more lateral visual areas, incorporating V1, V2 to V3. During the counting task compared to the memorization task, lower correlations were found in dorsal medial visual areas.
With these comparisons, it is theoretically possible to have significant effects for multiple comparisons in the same voxel (i.e. the correlation is significant in two conditions but in the opposite direction). However, we did not find overlap in any voxel at the employed statistical threshold. Therefore, each voxel shown here had a significant instructor-observer coupling only in a single condition.

2.3.4 Instructor-observer correlations across learning

Comparison between first and last learning run

Instructor-observer coupling changed was assessed across the learning runs by comparing the two extremes, the first and the third run. During the first run, all subjects started with a novel stimulus and attempted to encode all of the folding steps, whereas, in the last run, few to no steps were additionally learned since most subjects had already managed to almost fold the origami completely before the third run (see fig. 2.3). During the third run, we hypothesized that subjects monitored their previous performance instead of learning and postulated a stronger instructor-observer coupling in the first run. Indeed, we found significantly higher instructor-observer correlations in the first run than in the third run (see table 2.3 and figure 2.6) in cerebellar and visual areas. In the cerebellum the biggest clusters were in medial lobules crus 1 and 2 and on the left hand side in lobules VIIIa and b. The visual areas include inferior occipital areas and motion sensitive areas (V5, LOTC). Smaller clusters were also found in the precuneus and superior parietal lobule. No significantly higher correlations were found in the third run compared to the first run.For visualization purposes, we determined parameter estimates for each of the three learning runs in a repeated-measures ANOVA design. When inspecting the parameter estimates, we found a downward cascade of instructor-observer correlations over the course of the learning runs (see fig. 2.6c and d).

Relation between instructor-observer synchronization and performance in the premotor cortex

We hypothesized that the synchony between instructor and observer in the premotor cortex is related to the subsequent folding performance of the observer. We expected higher signal similarity to correspond to better performance. We chose the primary auditory cortex as a control region where we did not expect to see a relationship with performance. Since the performance data only shows reasonable variability in the first run (see fig. 2.3), we only assessed the this relationship in the first run. Our ROI analysis showed positive correlations in the premotor cortex. When Bonferroni-correcting for the number of ROIs, we found a significant correlation of 0.53 between the observer-instructor correlation and the performance in the left ventral premotor cortex which overlaps with BA 44. Positive correlation were found in all of the ROIs tested, including the auditory control ROIs, where the lowest correlations were found. The locations of the spherical ROIs and the resulting correlations are shown in Figure 2.7. To test whether the correlation with performance is specific to learning, we correlated the synchrony during fold counting in the premotor ROIs with the performance of the first learning run. No significant correlation was found (left vPMC ρ=-0.01, right vPMC ρ=0.06, right dPMC

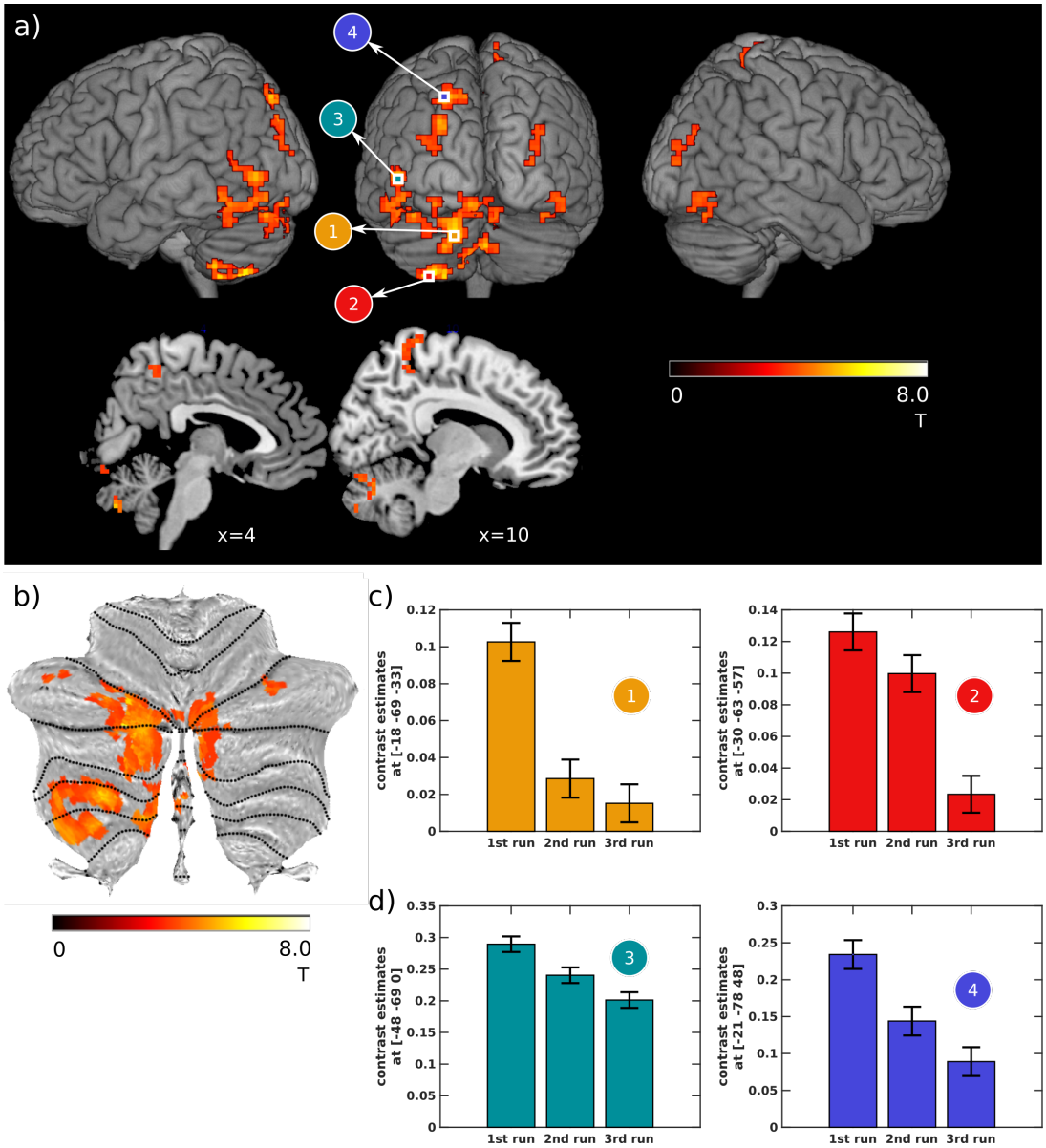


Figure 2.6: Instructor-observer synchrony decreases across learning runs. The synchrony declines between first and last learning run in visual, parietal and cerebellar areas. The t-statistic is thresholded at T_{thr}=3.42. a) an overlay of significant differences between the first and the last run. Only positive correlations with 'first > last' reach significance. b) cerebellar flatmap depicting higher correlation for memorization than for counting run. Equivalent to cerebellar part of the heatmap in a). c) and d) contrast estimates for exemplary voxels, errorbars denote the standard error of the mean, in cerebellar voxels (c) and in cortical voxels (d).

location	cluster size	peak statistics x	y	z	T	peak location *area with max probability*
first watch to memorize > third watch to memorize ∩ first watch to memorize > 0 (positive correlations)						
R+L Cerebellum; Lingual Gyrus	221	-18	-69	-33	6.07	L Cerebellum (Crus 1); Lobule VIIa crus I (Hem) 28%
		-15	-78	-33	5.89	L Cerebellum (Crus 2); Lobule VIIa crus I (Hem) 77%
		-21	-69	-39	5.06	L Cerebellum (VIII)
		-6	-87	-15	4.89	L Lingual Gyrus; Area hOc2 [V2] 32%
		-39	-75	-21	4.86	L Cerebellum (Crus 1); Area hOc4v [V4(v)] 48%
		-27	-84	-15	4.57	L Lingual Gyrus; Area hOc4v [V4(v)] 82%
		9	-69	-24	4.48	R Cerebellum (VI); Lobule VI (Hem) 49%
		9	-69	-30	4.41	R Cerebellum (VIII)
		9	-78	-21	4.17	R Cerebellum (VI); Lobule VI (Hem) 95%
L Inferior Temporal Gyrus; Occipital Cortex	134	-48	-69	0	5.12	L Middle Occipital Gyrus; Area hOc5 [V5/MT] 49%
		-51	-54	-15	4.56	L Inferior Temporal Gyrus
		-45	-75	-12	4.48	L Inferior Occipital Gyrus; Area hOc4la 65%
		-48	-51	-12	4.46	L Inferior Temporal Gyrus; Area FG4 38%
		-51	-66	-15	4.43	L Inferior Occipital Gyrus; Area FG2 34%
		-48	-69	-12	4.42	L Inferior Occipital Gyrus; Area FG2 39%
		-54	-51	-12	4.42	L Inferior Temporal Gyrus
		-48	-51	-21	4.32	L Inferior Temporal Gyrus; Area FG4 45%
L Cerebellum (VIII)	100	-30	-63	-57	6.29	L Cerebellum (VIII); Lobule VIIIa (Hem) 78%
		-21	-48	-57	5.56	L Cerebellum (VIII); Lobule VIIIb (Hem) 94%
		-33	-45	-51	5.31	L Cerebellum (VIII); Lobule VIIIa (Hem) 71%
		-24	-42	-48	4.96	L Cerebellum (VIII); Lobule VIIIb (Hem) 62%
		-33	-51	-51	4.19	L Cerebellum (VIII); Lobule VIIIa (Hem) 64%
R+L Cerebellum	52	3	-75	-39	5.30	Cerebellar Vermis (8); Lobule VIIIa (Verm) 85%
		-9	-66	-51	5.27	N/A; L Lobule VIIIb (Hem) 60%
		-6	-66	-45	4.80	L Cerebellum(VIII); Lobule VIIIb (Verm) 69%
		0	-72	-42	4.78	Cerebellar Vermis (8); R Lobule VIIIa (Verm) 88%
		9	-81	-42	3.95	R Cerebellum (Crus 2); Lobule VIIa crusII (Hem) 54%
		-9	-78	-45	3.85	L Cerebellum (VII); Lobule VIIb (Hem) 70%
L Occipital Cortex	44	-24	-81	33	5.02	L Superior Occipital Gyrus
		-33	-84	21	3.95	L Middle Occipital Gyrus
		-27	-87	18	3.89	L Middle Occipital Gyrus; Area hOc4lp 34%
R Occipital Cortex	43	36	-69	-21	4.31	R Cerebellum (VI); Area FG1 53%
		39	-69	-9	4.19	R Inferior Occipital Gyrus; Area FG1 33%
		45	-72	-12	4.16	R Inferior Occipital Gyrus; Area hOc4la 60%
		48	-75	-18	4.09	N/A; R Area hOc4la 51%
		45	-66	-15	3.88	R Inferior Occipital Gyrus; Area FG2 71%
R Precuneus	41	6	-45	57	4.48	R Precuneus; Area 5M (SPL) 56%
		9	-39	75	4.27	R Paracentral Lobule; Area 4a 41%
		9	-48	69	4.07	R Precuneus; Area 3b 36%
		12	-51	57	3.99	R Precuneus; Area 5L (SPL) 35%
L Superior Parietal Lobule	30	-21	-78	48	5.31	L Superior Parietal Lobule
		-9	-81	48	4.49	L Precuneus; Area 7P (SPL) 60%
R Middle Occipital Gyrus	26	30	-84	9	4.10	R Middle Occipital Gyrus
		36	-75	27	3.97	R Middle Occipital Gyrus
		33	-87	18	3.89	R Middle Occipital Gyrus
third memorize > first memorize ∩ third memorize > 0 (positive correlations)						
n.s.						
third memorize < first memorize ∩ third watch to memorize < 0 (negative correlations)						
n.s.						
first memorize < third memorize ∩ first memorize < 0 (negative correlations)						
n.s.						

Table 2.3: Peak voxels and brain regions with higher synchrony in the first compared to the third learning run. This table corresponds to figure 2.6. The statistical procedure is described in materials and methods. Significantly higher correlations were only found in the first run, not in the last run.

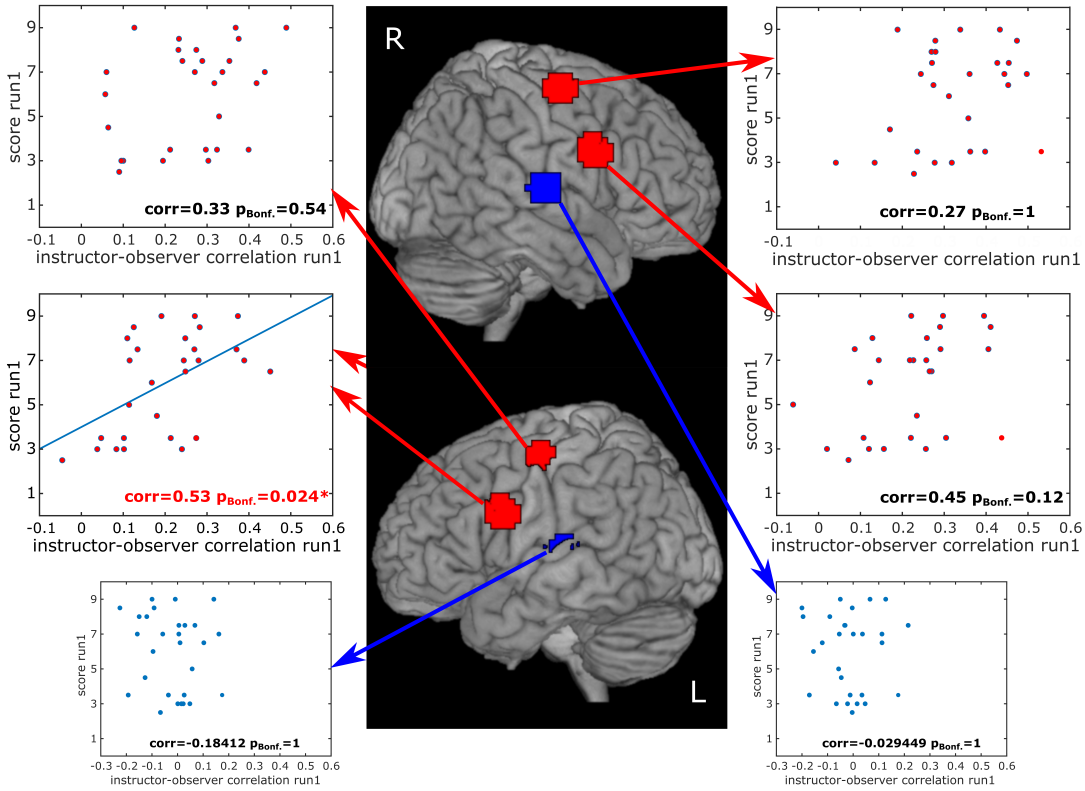


Figure 2.7: Left ventral premotor activity correlates with behavior. Spearman rank correlations within predefined ROIs in the first learning run. A significant correlation between instructor-observer-synchrony and subsequent performance was found in the left ventral premotor cortex. The asterisk denotes that this correlation fulfills p<0.05 Bonferroni-corrected for the number of ROIs. The lower two graphs show the two auditory control ROIS were no significant correlation was found. Note that the linear regression line is shown for display reasons only as the Spearman correlation captures the strength of a monotonic relation instead of the goodness of a linear fit.

ρ=-0.24, left dPMC ρ=-0.08). The significant correlation in the left vPMC is significantly higher than in the count folds run (p=0.017, one-sided). For exploratory purposes we also calculated the correlation between synchrony and performance in a voxel-wise whole-brain analysis. Although a number of brain regions show high correlations, at a statistical threshold of $q_{FDR}<0.05$ or $p_{uncorr}<0.001$ no cluster of reasonable size survives. The results can be found in the supplementary information figure 2.11.

2.3.5 Temporal shifts

We also wished to investigate temporal shifts in the correlation between instructor and observers. To this end, we characterized the lag/delay structure for all of the voxels in the brain. The distribution of median lags was centered around lag '0' for all subjects and voxels within the whole grey matter mask (fig. 2.8a). We therefore conclude that no noteworthy delay exists between the observers and instructor brain time courses, i.e. that they are in sync without any lag. Still it is possible that individual brain regions show a systematic lag or lead. We tested for a significant difference from '0' with a two-sided Wilcoxon signed-rank test and found two significant clusters for the first learning run. Data are shown in Fig. 2.8b-d. The clusters reside in the postcentral gyrus and inferior parietal lobule. For both clusters the observer activity led the instructor activity. None of the other conditions showed a significant delay at this statistical threshold.
To see whether this effect is specific to the first learning run, we compared the delays between the first and the last learning run as well as between the first learning run and the control condition of counting folds. We found only small clusters using a two-sided Wilcoxon signed-rank test on the lag /delay differences ($p_{uncorr.} < 0.001$). The direction of the lag differences suggest that this effect is specific to the first condition. For the contrast 'first memorize' vs. 'count' we found a voxel count k=12 and peak lag difference=6; and k=8, peak lag difference=7.5. For the contrast 'first memorize' vs. 'third memorize' we found k=10, peak lag difference=9; and k=5, peak lag difference=6.5. Lag differences were defined as $lag_{\text{first memorize run}}$-$lag_{\text{comparison run}}$, such that positive delay differences correspond to more positive delays in the first memorize condition. In other words, the observer advances more in the first run.

location	cluster size	peak statistics x	y	z	p	delay	peak location *area with max probability*
R Postcentral Gyrus, R Inferior Parietal Lobule	58	51	-21	36	4.0e-6	4	R Postcentral Gyrus; Area PFt (IPL)
		51	-21	42	7.0e-6	4	R Postcentral Gyrus; Area 1
		63	-24	45	7.0e-6	8	R SupraMarginal Gyrus; Area PFt (IPL)
		39	-27	45	9.1e-5	5	R Postcentral Gyrus; Area 3b
L Postcentral Gyrus, L Inferior Parietal Lobule	44	-45	-30	42	4.0e-6	6.5	L Inferior Parietal Lobule; Area PFt
		-57	-24	36	3.9e-5	6	L SupraMarginal Gyrus; Area PFt (IPL)
		-60	-27	33	4.4e-5	6	L SupraMarginal Gyrus; Area PFt (IPL)

Table 2.4: Peak activity locations for clusters with significant lags/delays of the first memorize run at $p<0.001$ uncorr., $k \geq 24$. P-values were determined by Wilcoxon signrank test. The delay is given in multiples of TR with 1 TR=0.59 s. A positive delay means the observer precedes the instructor.

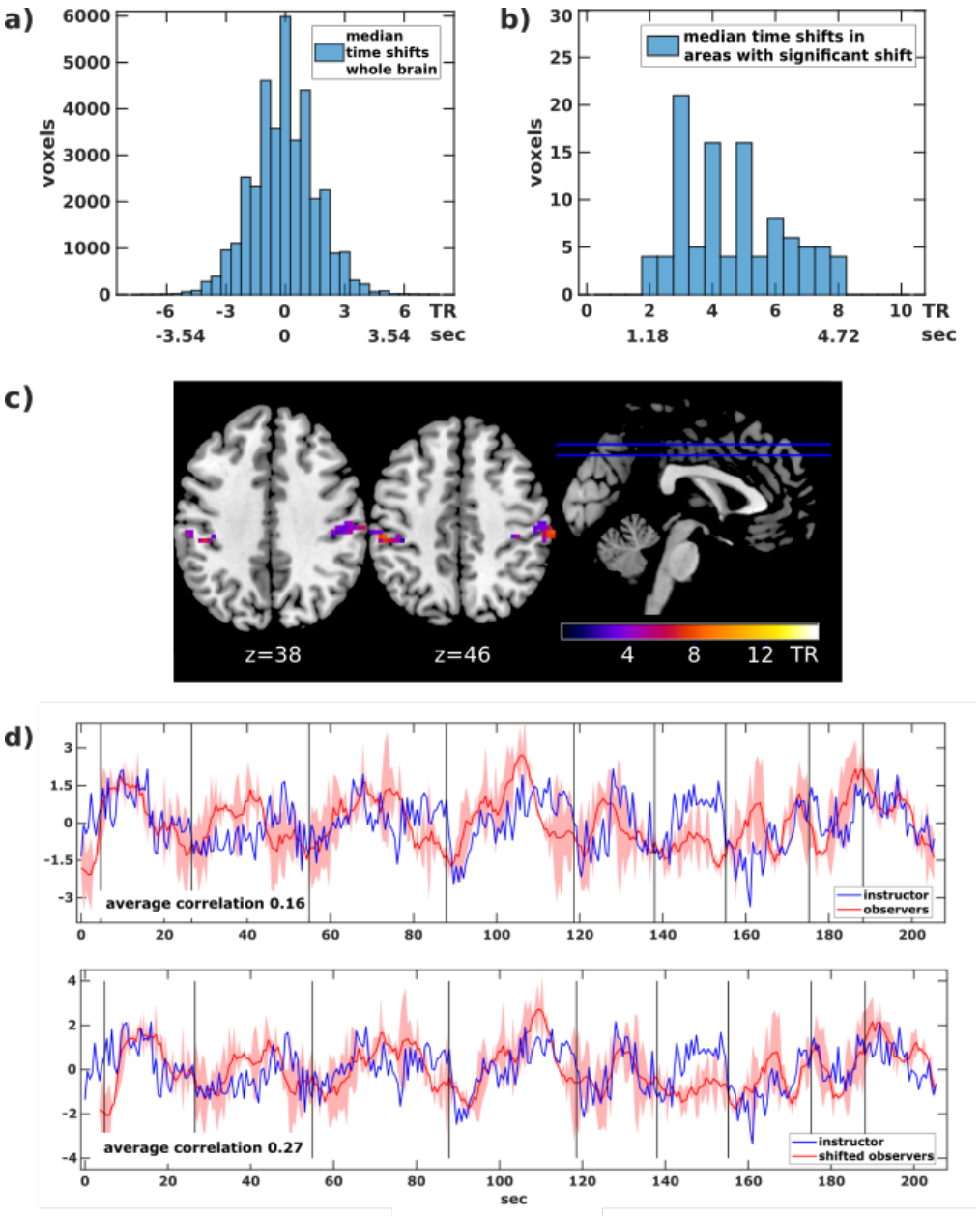


Figure 2.8: Temporal shifts between instructor and observers. Most voxels show no significant delay between instructor and observer (first learning run). However the observer activity significantly advances the instructor activity in the anterior IPS during the first learning run. A positive delay means observer advances. a) histogram of the distribution of the median delays over whole grey matter mask (median over all subjects in that voxel). b) histogram of median delays for voxels with $p<0.001$ uncorr. and $k\geq 24$. c) locations of the voxels with significant median delays. d) time course of an exemplary significant voxel (-45, -30, 42, median delay 6.5·TR or 3.84 sec). The upper image shows the two time courses at lag '0', the lower one with a shifted observer time course to match the lag. The red solid line is the average observer time course, the red shaded area corresponds to one standard deviation from the mean.

2.4 Discussion

This study aimed to investigate the similarity in brain activity between an instructor and observers while the observers learn a new skill; in our case to fold an origami. We demonstrate a synchrony in the BOLD signal between instructor and observer throughout the AEON, for the first time in the absence of a common sensory modality, thereby expanding the classical GLM-based literature on the AOEN. This synchrony does not appear to depend on the task, but is rather present in all of the conditions we tested. The cerebellum, on the other hand, plays an differential role in instructor-observer-coupling depending on the action observation task. The cerebellum has been a strongly neglected brain area in action observation research until recently. In the ventral premotor cortex, the strength of this synchrony correlated with behavioral performance. Lastly, we present a bootstrap routine that explores the temporal relationship between observers and instructor. These results suggest that most of the brain shows no lag between the instructor and observer, except for an area in the postcentral gyrus and inferior parietal lobe, where the observers' activity advances the instructor's activity.

2.4.1 Overall observer-instructor synchronization

Although a vast literature exists on the overlap in brain activation between action observation and action execution, few studies have investigated the temporal similarity of these processes in humans (see Mukamel et al. (2010) as an exception). Therefore, one study aim was to show that activity during action observation is synchronized to the activity during action performance, and to rule out that alignment can be explained by common visual processing. For this reason, the instructor was blindfolded, performing her actions solely with somatosensory and proprioceptive feedback. The observers also watched the motor action without moving. Therefore, no task-related sensory or motor activity was similar between the instructor and the observers. Instead the synchrony we demonstrate is between motor and visual processing. Athough overall our results match magnitude based activation/deactivation studies (Pajula et al., 2012), these methods investigate different aspects of the BOLD signal (see the appendix of Thomas et al. (2018) for a comparison between the two methods).

The classical bilateral AOEN network was synchronized in all the cognitive tasks that were tested (learning, monitoring and counting). These areas include the dorsal and ventral premotor cortex, the somatosensory cortex, the superior and inferior parietal lobe as well as area V5 (see the meta-analyses of Caspers et al. (2010) and Hardwick et al. (2018) for a list of the AOEN regions). Brain areas involved in movement execution were also synchronized between instructor and observers including the premotor and somatosensory areas as well as the cerebellar lobules VI and VIIIa and b (Hardwick et al., 2018;

Schlerf et al., 2010). Interestingly, the SMA did not show consistent synchrony across all conditions.

Our temporal similarity analysis can be regarded as an analog and complementary to a classical conjunction analysis between movement observation and execution (except that in our case observation and execution are not performed by the same subject). The cortical areas we find match the the meta-analyses of the conjunction of movement execution and action observation well (Caspers et al., 2010; Hardwick et al., 2018). We additionally find the right TPJ as well as a cluster on the border between BA45 and the DLPFC. Similar areas were reported in studies about immediate imitation (Buccino et al., 2004; Vogt et al., 2007; Higuchi et al., 2012; Sakreida et al., 2017) and using intersubject-correlation for subjects observing action sequences Thomas et al. (2018).

Data on action observation in the cerebellum are the sparse and partly contradicting. In the meta-analyses of action observation (Caspers et al. (2010); Hardwick et al. (2018)), cerebellar activity is not significant during action observation and therefore also not in the conjunction analysis of action observation and movement execution. However, in a meta-analysis of the cerebellum, extensive clusters for 'mirroring' , which includes action observation, were found (Van Overwalle et al. (2014)). Two recent studies have found the cerebellum in classical GLM-based tasks. Abdelgabar et al. (2018) found the cerebellum using an extensive dataset (79 participants) on action observation and Casiraghi et al. (2019) found cerebellar activity with an action execution and observation grip force task. Our conjunction results show a good overlap with the first study (Abdelgabar et al. (2018)), with cerebellar clusters in the bilateral lobules VI and VIIIa and VIIIb. The pattern of cerebellar synchrony we found in the first learning session also overlaps with the AOEN network from the second sudy (Casiraghi et al., 2019). Studies on immediate imitation have also reported cerebellar activation in a conjunction between observation and manual execution (Vogt et al., 2007; Sakreida et al., 2017), where activity is found in VI but also VIIIa. The superior cerebellar clusters we found in our study overlap with reported cerebellar areas of motor execution of complex hand-finger movements (Schlerf et al., 2010; Mottolese et al., 2013). Taken together, the cerebellum appears to play an integral role in action observation and execution. The synchrony between observation and execution occurs in areas adjacent to the somatotopical areas for complex manual action (Schlerf et al., 2010; Mottolese et al., 2013). This hints to a similar processing of observation and action in the cerebellum, as has already been suggested for the cerebral cortex AOEN areas in context of the mirror neuron research (i.e. Rizzolatti and Craighero, 2004).

However, for different conditions of this experiment, we find the correlated clusters extending beyond the hand areas (e.g. crus 1 and 2), suggesting also a processing similarity of action and action obervation beyond hand execution areas.

Moreover, we found temporal similarities between instructor and observers in visual areas, also early visual areas (V1, V2, V3). Considering that the instructor is blindfolded and cannot see her hands or the paper, this poses an interesting side-finding. Visual activity is normally not expected during movement execution and visual feedback can be ruled out. An intuitive guess is motor imagery performed by the instructor when folding; however, at least on a meta-analyis level, the early visual cortex is not reported to be active in motor imagery (Hardwick et al., 2018). Nevertheless, as this study investigated similarities, results cannot be straightforwardly carried over from magnitude-based GLM findings (see explanation above). The visual cortex has been reported to be involved in sensory processing beyond visual input: Merabet et al. (2007) have found primary visual activation when processing tactile stimuli without visual input and in the mouse visual cortex, motor-related activity without visual input has been reported (Keller et al., 2012). Consequently, our results can be interpreted as a further evidence for non-visual sensory processing in primary visual areas (here in the case of the instructor). Moreover, the visual cortex activity is thought to be strongly related to predictions (Keller et al., 2012). For humans, a model relating visual cortex acitivity to expected input was proposed (Brandt et al., 2005). Consequently, these shared activity structures between blindfolded instructor and observers could also originate in expected visual input of the instructor during her motor actions. Consequently, these interesting findings should encourage further research on the relation of visual cortex activity and motor output. One more thing should be noted when integrating the results: an appealing aspect of investigating correlations is the fact that they are effect sizes. While the correlations in classical AOEN areas are fairly high also when comparing to other ISC studies (z-values between 0.3 and 0.5), the correlations in visual areas (excluding V5/MT) are fairly low, although still typical for ISC studies (z-values (fisher) around 0.10 and 0.15) (for an ISC study reporting average correlations per voxel, see e.g. Herbec et al., 2015).
Overall, the findings demonstrate a common coupling of instructor and observers during different tasks of action identification (learning, action monitoring and counting) over a broad range of areas. Consequently, a similar processing between action performance and action observation on a voxel/cluster level is to be assumed. We can now expand this knowledge on common activation to temporal similarities, showing a good match with recent findings on action observation in the cerebellum, and also expanding these commonalities found in action-execution conjunctions to a non-imitation task (counting folds).

2.4.2 Memorizing/encoding vs. identifying actions

We found higher instructor-observer correlations for memorizing than for counting folds especially in cerebellar areas, but also in lower visual areas, i.e. primary visual areas.

In classical AOEN areas we found the left V5 and a small cluster in the inferior parietal lobe, both classical motion and spatial processing areas. Other classical action observation areas like the somatosensory cortex, or dorsal or ventral premotor cortex did not show significant differences at our employed statistical threshold. This might be intriguing given the fact that these areas are often discussed for 'action understanding' (e.g. Rizzolatti and Craighero, 2004). However, in contrast to other experiments, our control condition also involved action understanding as one needs to understand and identify what a fold is, in order to count one. Moreover, not much can be inferred from a non-significant result.

The higher instructor-observer coupling is mainly found in cerebellar crus 1 and 2, but also in adjacent parts of lobules VI and VIIa. The latter can be described in terms of higher sensorimotor similarity to the instructor during observation as the observers need to map the observed action onto their own motor repertoire during learning for the subsequent reproduction.

The interpretation of crus 1 and 2 however is interesting as these areas are classically not described as motor regions of the hand. Nevertheless, our data show a good overlap with the observational weight discrimination task of Abdelgabar et al. (2018); the posterior crus 1 and 2 activations also overlap with and are adjacent to the mirroring meta-analysis of Van Overwalle et al. (2014). Consequently, these areas have been described in the context of action observation even though not being primary sensorimotor areas.

In mice, Proville et al. (2014) show reciprocal connections between crus 1 and M1 as well as S1 for whisker control. They conclude that the cerebellum combines sensory and motor information for 'a deeper understanding of the sensory processes'. Along with this finding, Ronconi et al. (2016) argue that the cerebellum is vital for multisensory integration. In this sense, our higher cerebellar similarities for learning than counting might reflect a sensorimotor integration more similar to the instructor when memorizing the stimulus in order to reproduce it.

Another non-exclusive interpretation of the findings is a difference in cognitive control. If we compare our results to the recent network analysis of Marek et al. (2018) we find a good overlap with mainly the fronto-parietal network, but also to a lower extent with the cingulo-opercular network and the hand-sensorimotor network. The first two are intriguing as they are associated with cognitive control, executive function and sustained attention. Cortically, a multiple-demand system has been proposed (Duncan, 2010) which modulates with task difficulty (Fedorenko et al., 2013) and shows strong overlap with the fronto-parietal and cingulo-opercular network (Camilleri et al., 2018). Indeed, it can be assumed that while both observer tasks were designed to ensure attentiveness and cognitive processing, the task difficulty for learning was higher than for counting.

Consequently, a higher activity due to task difficulty could be associated with higher synchrony to the instructor. However, it should be noted that we do not find a matching cortial pattern at our employed statistical threshold.
The instructor-observer correlations in crus 1 and 2 show a good overlap with clusters previously associated with visuo-spatial working memory (Stoodley et al., 2012; Küper et al., 2016). Thus, an explanation for the higher instructor-observer coupling in cerebellar areas is not only a 'motor-coupling' in the corresponding cerebellar motor areas, but also similar processes of spatial working memory between instructor and observers: It should be assumed that the instructor utilized similar spatial working memory for the paper folding sequences of both the butterfly origami and the unproductive sequence given also the fact that both sequences were trained. However, it is reasonable to assume that the cognitive load on spatial working memory is higher for learning than for counting. This higher load for the observers could translate into a higher similarity to the instructor's spatial working memory processes, leading to higher instructor-observer-synchrony during the learning session. (See also Küper et al., 2016 for a report of higher cerebellar involvement with a higher cognitive demand of abstract memory.)

2.4.3 Synchrony reduces during repeated viewing

Between the three learning runs we compare the two extremes, i.e. the first and the last run. All but four subjects (i.e. 24 subjects) have almost or totally managed to fold the figure after the second run (see fig. 2.3. We therefore assume that the contrast between the first and last run shows the contrast between encoding a new stimulus vs. the observation of a known stimulus presented to update and compare the observers' own motor repertoire to. No voxel of the third run is showing a significantly higher correlation than in the first run, suggesting an overall loss of synchrony between instructor and observers. This loss is happening mainly in non-classical AOEN areas. This suggests that especially cognitive processes that are supportive to the AOEN function like visuo-spatial working memory are reflected in these synchrony changes.
Here, we cannot rule out that the findings are a pure attention effect - observers pay lesser attention to the incoming stimulus and therefore how less synchrony to the (attentive) instructor. Nevertheless it is important to state that the aim of this study was to perform a naturalistic experiment. It is to be assumed that attentiveness and cognitive load decline during repeated viewing, as more parts of the video are encoded. Consequently, this is a part of naturalistic learning.
Another possible explanation for the seen reduction of synchrony are repetition suppression effects (see e.g. Grill-Spector et al., 2006). A lower observer-instructor correlation could originate from a reduced observer signal due to higher relative noise. Grill-Spector et al. (2006) propose several mechanisms for a seen signal reduction due to repeated

stimulus. Apart from a neuronal fatigue model, neuronal sharpening (i.e. fewer neurons involved), or neuronal facilitation (improved prediction due to repetition (Friston, 2005)) are discussed. Consequently, it is possible that the learning process resulted in a more efficient stimulus processing.
Significantly higher instructor-observer-correlations in the first run are found in cerebellar regions of posterior crus 1 and 2 and lobules VIIIa and VIIIb, as well as occipital and posterior parietal clusters, plus a precuneus/SPL cluster (BA5). The crus 1 and 2 clusters overlap with those of the memorizing vs counting contrast. Analog to the memorizing vs. counting contrast, we can also assume for these clusters that a stronger coupling implies a more similar visuo-spatial processing and working memory between instructor and observer during the encoding of the motor actions than during the comparison of own motor repertoire to the instructive video. Moreover, the VIIIa and VIIIb cluster overlaps with the cingulo-opercular control network and the premotor network of Marek et al. (2018) which could indicate less motor-encoding.
The cerebral cortical similarities could follow an explanation analog to the cerebellar similarities: higher synchrony can reflect higher similarity of the observer to the instructor in spatial processing and in sustained attention (SPL/precuneus), in bottom-up attention as represented by the dorsal attention network (occipito-parietal clusters) and/or in the higher need for multisensory integration during the initial learning stage. Our findings are also consistent with the reportings of other immediate guitar chord imitation studies, which have found significantly higher activity during action observation in AOEN areas as well as similar or adjacent cerebellar areas for non-practiced compared to practiced chords. But note that for these studies, a practiced motor action (the pressing of a guitar chord) means that this motor action is consolditated for at least one day (Vogt et al., 2007; Higuchi et al., 2012).

2.4.4 Better reproduction is associated with higher instructor-observer correlation in the left ventral premotor cortex

The premotor cortex plays a vast role in motor control, planning and learning (see e.g. Kantak et al., 2012 for a review). Especially the ventral premotor cortex is a classical area of the AOEN as in area F5 the first mirror neurons have been found (see e.g. Rizzolatti and Craighero, 2004 for a review). Area F5 in non-human primates is thought to be homolog to the ventral premotor cortex and BA44 in humans (Binkofski and Buccino, 2006). Classically, this area has been associated with 'action understanding' in the mirror neuron literature (see e.g. Rizzolatti et al., 2002). Specifically for the left ventral premotor cortex, involvement in egocentric finger action imagery (Binkofski et al., 2000) or modulation with difficulty for object-hand posture mapping has been reported

(Vingerhoets et al., 2013).
For our ROI analysis of the premotor cortex we found that subjects who performed better subsequently tended to have higher correlation to the instructor activity in the left ventral premotor cortex. This result fits well to the recent findings of Pan et al. (2018) who report a relation between learning success in a teacher-learner interaction when learning a song in the inferior frontal cortex. However, learning a song and learning an origami are not completely comparable as the song also has a language component. While language is also left-lateralized and Broca's area also incorporates BA44, this 'language location' of BA 44 has been reported to be adjacent to the 'action observation execution' location (Cerri et al., 2015). A very recent study by (Sacheli et al., 2019) investigated interacting with a (virtual) partner using a music box and found involvement of the left vPMC exclusively during interaction; moreover, the left vPMC activity decoded the actions of the partner.
Apart from this, teacher-learner situations in the context of observational skill/motor learning remain sparsely investigated. Nevertheless, there is evidence that the premotor cortex activity modulates with learning/training (see Cross et al., 2009; Higuchi et al., 2012; Wiestler and Diedrichsen, 2013; Kirsch and Cross, 2015; Apšvalka et al., 2018). Moreover, transcranial direct current stimulation (tDCS) over the premotor cortex during action observation of keyboard press sequences leads to better reaction times in the reproduction phase (Wade and Hammond, 2015). Kirsch and Cross (2015) also report a correlation between left PMC activity and behavioral scores after one week of training dance sequences.
Taken together, the left vPMC modulates with training, is specifically active during interaction and decodes the actions of the partner. We show that a coupling between observer and instructor can be associated with better subsequent performance in this area. Kilner (2011) argues that the vPMC is responsible for direct action selection from goal to performance and to sensory consequences. Or resuts suggest that if this concrete mapping onto own motor repertoire is more similar to the instructor, a more similar (and therefore better) reproduction is possible. In line with the newest research by Sacheli et al. (2019) a higher similarity to the instructor would mean a better decoding of instructor's actions, and a better decoding would mean a mapping onto own motor repertoire that is more similar to the instructor's mapping to actions. This would lead the observer to a better reproduction.
Our whole-brain analysis did not yield significant results. However, it depicts an unterlying pattern of correlation. It shows that also other areas of the premotor cortex show high correlations with performance. These peaks are adjacent to the placement of the ROIs generated from the meta analysis by Hardwick et al. (2018). Therefore, the data suggests that also other parts of the premotor cortex are involved in this process but the

results remain inconclusive. We encourage further research.

2.4.5 Exploring the temporal relations

One aim of the study was to investigate temporal relations given the findings in previous studies on spoken interaction (Stephens et al., 2010; Zadbood et al., 2017; Piazza et al., 2018) and the high sampling rate compared to previous fMRI studies. For most of our voxels and conditions, instructor and observers are in sync without lag. Nevertheless, we also find a bilateral cluster in which the observer signal is advancing the instructor signal at a statistical threshold of $p_{uncorr.}<0.001$. In contrast to Stephens et al. (2010) or Piazza et al. (2018), we do not find anticipatory signals in the 'receivers' (in their case listeners, in our case observers) in prefrontal areas. (But note the methodological differences, see materials and methods.)
We observe two bilateral clusters in the postcentral gyrus and inferior parietal lobule for the first memorize run; closer visual inspection suggests these clusters to reside alongside the postcentral sulcus, overlapping with the anterior intraparietal sulcus (aIPS). This is a 'classical' AOEN resp. mirror neuron system area, which is active during performing a manual action and observing a manual action(Rizzolatti and Sinigaglia, 2010). Hamilton and Grafton (2006) reported this area to represent 'goals' of observed actions (immediate goals like 'take a cookie'). The inferior postcentral sulcus and anterior intraparietal sulcus have also been reported to be related to the learning of continuous motor sequences (Reithler et al., 2010). This constitutes a border area; while the inferior parietal lobule is involved in different spatial operations, the postcentral gyrus incorporates the somatosensory areas. Tanabe et al. (2005) have found a very similar cluster to be responsible for crossmodal integration of stimuli. For other conditions, we do not find a significant delay; data suggests this delay may only be present in the first learning run given that testing to compare between the learning runs shows a trend for this. However, as these thresholds are low, we can only argue for a tendency which needs to be confirmed. Consequently, one possible explanation might be a cognitive anticipation during learning whereas the instructor has to integrate the perceptive signals. For the instructor, we can assume that the strokings of the folds which happen at the end of the step are more important due to the blindfold. Consequently, spatio-tactile information might be integrated later in the instructor's aIPS than action mapping in the observer. Another possible explanation would be alonside the findings of Hamilton and Grafton (2006): In the inital learning condition, in order to map the actions onto their own motor repertoire, it may be favorable for the observers to anticipate the immediate goal (i.e. the end product of the folding) of the movements.
Note that these findings, while intriguing, have to be handled with care, given the statistical thresholding. Nevertheless, we believe this is a valuable contribution given that

previous interaction studies have reported possible anticipation. We suggest to take this result as an incentive for further investigation.

2.4.6 Cerebellar contributions

It is intriguing that a great part of the findings of this paper are located in cerebellar areas. In summary, classical hand sensorimotor areas are consistently synchronized between instructor and observer, whether the task is learning or counting. On the other hand, changes in instructor-observer-synchronization between these cognitive tasks reside in cerebellar areas that are not classically associated with hand motion. Based on the current literature a likely interpretation is a change in similarity between the instructor and observer in processes of attention, spatial working memory and cognitive control. This would also be consistent with the theory of the cerebellum being especially involved in multisensory integration during the initial stage of learning when visual input needs to be mapped onto own motor repertoire.

Nevertheless, the current fMRI literature on the cerebellum is sparse (Fair, 2018), especially considering action observation Abdelgabar et al. (2018). The reasons for this are manifold: in former times, to trade off between spatial coverage and spatial and temporal resolution, often only the superior parts of the cerebellum were imaged, and up to version 8 SPM was also utilizing a default bounding box which does not cover the inferior parts (Abdelgabar et al., 2018). Moreover the signal-to-noise-ratio is also lower for cerebellar than for cortical signals leading to less consistent results (Fair, 2018). We also experience lower correlations in the cerebellum than e.g. in somatosensory or premotor areas; Abdelgabar et al. (2018) also reports 'slightly less reliably' recruited cerebellar areas compared to certain classical cerebral action observation ROIs, in terms of comparing the number of participants showing significant activity.

In our case, the differences between the sessions range mainly in cerebellar areas, but also with smaller extent in 'classical' AOEN areas. Given the scant literature on cerebellar activations during action observation, the interpretations of the results have to remain partly speculative. (Note that while motor or sequence learning is known to be partly mediated by the cerebellum, it is important to state that in those studies normally visual cues are followed by motor output; in our case, we only compare pure observation with different cognitive tasks but without parallel execution to motor execution without observation.)

Results from the midnight scan club, an initiative to aquire extensive datesets of single subjects, suggest a high inter-personal variability for brain networks, especially for the cerebellum (Gordon et al., 2017; Marek et al., 2018). Even though the present paper does not analyze networks within subjects, we can still state that the cerebellar locations found in this paper have an uncertainty when inferring population results as we only

have one instructor. On the other hand, the fact that we have one instructor might have helped us to investigate the cerebellar involvement given that several instructors may have 'smeared out' the activation.

2.4.7 Limitations and outlook

Needless to say, this study has also certain limitations: first and foremost, this experiment has only one instructor. To show every subject the same video for exact comparison, this is a necessity.

Silbert et al. (2014) have expanded the speaker-listener study of Stephens et al. (2010) by scanning three subjects who have vastly practiced the original story to be able to tell it reproducably inside the scanner. By dilating and stretching the time course based on the audio track (time warping the signal), the subjects could be mapped onto the original speaker. This provides a great option to show the reliablity of the speakers activity. However, it should be considered that in this case the subjects still experience only this one story. Moreover, given the high variability of cerebellar networks between subjects (Marek et al., 2018) and the high signal-to-noise ratio, it might be sensible to consider several sessions of one instructor when the aim is to improve reliability of the results.

In case it is not a necessity to observe the same video, the setup could and should be expanded to a complete dyadic setup as used in e.g. Schippers et al. (2010): to create pairs of instructor and observer. We would like to remark that the comparability between dyads will be less stringent and the stimuli less controlled due to variations. While this dyadic setup requires a larger number of subjects, it would be a great approach to i.e. investigate the temporal delay structure across different instructor-observer performance-couplings. It should be noted that while hyperscanning offers great possiblities for this, in case of a simple instruction-observation pairing without online performance feedback, no hyperscanning is needed.

Furthermore, it is to be noted that the instructor was trained. Our aim was to reconstruct a naturalistic instructor-learner situation; therefore, a trained instructor is true-to-life. Moreover, the training was necessary for our setup so that the instructor was able to fold the origami in a naturalistic and faultless way when blindfolded. For the future, it would be intriguing to investigate different instructor levels. This could mean either a badly-performing instructor, or a different training level: it would be interesting to show the video of either a person who has freshly managed to fold the origami or that of an origami expert to a learner and look for similarities and differences between these subjects. It is known that trained sequences can lead to less BOLD signal change (see e.g. Reithler et al., 2010; Wiestler and Diedrichsen, 2013) and observation of novices engages parts of the AOEN more strongly than observation of experts (Errante and Fogassi,

2019). However, on the other hand, the later study also showed a relation between activation in parts of the AOEN and own manual abilities specifically when observing the expert. Finally, one cannot directly infer the strength of intersubject similarity from stronger activation. (See the supplemental information of Thomas et al. (2018) for a comparison between the general linear model and the intersubject correlation approach.) Consequently, pairing different instructor-observer performance levels can lead to further insights into the instructive learning process.

2.4.8 Conclusion

With this paper, we expand the classical GLM literature on the AOEN to a common temporal structure between action and observation. Additionally, we show that a better coupling to the instructor during observation tends to be associated with better performance afterwards. Moreover, we show instructor-observer-coupling differences between different cognitive observation tasks to reside to a great part in cerebellar areas, which is a not well-studied and a neglected part of the AOEN. As we see the greatest effects in the cerebellum, we strongly recommend to investigate cerebellar aspects to observational learning in the future. New imaging options (multiband imaging) offer a better spatial and temporal resolution for this. The better temporal resolution is intriguing for more fine-grained temporal analysis. We have made a suggestion how to approach this in this paper and would encourage other readers to further this research and methods.
With the presented paradigm one can investigate interactive observational learning as a type of social interaction. Since social interaction is a neglected but important neuroscientifc field of research, also for neuropsychiatry (Schilbach, 2016), it will be important to further paradigms similar to this one, and to adapt it to different subject and patient groups.

2.5 Supplementary material

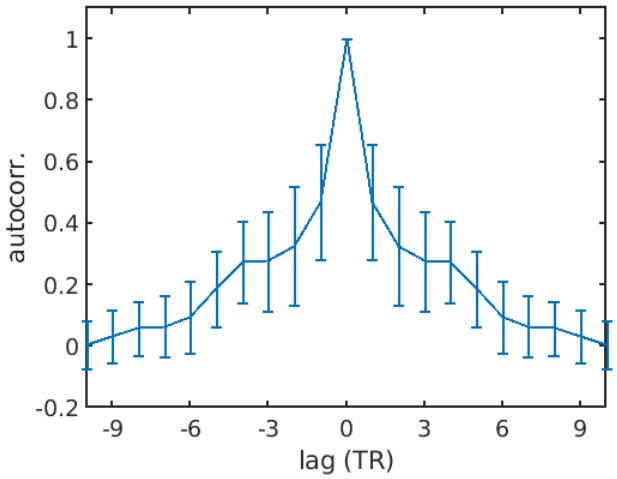


Figure 2.9: Autocorrelation of the instructor image for the butterfly condition. Mean autocorrelation within the grey matter mask plotted, errorbars denote the standard deviation.

Table 2.5: The questionnaire evaluating whether subjects suspected the blindfold:
Please answer with yes or no:
While you were watching the instructive videos, did you at any point think that:

the two hands belonged to two different actors	yes	no
the person in the instructional video changes between the videos	yes	no
the person in the video could not see what he/she was doing	yes	no
the person in the video could not feel what he/she was doing	yes	no
the person in the video received constant auditory feedback	yes	no
the origami paper was different to your paper	yes	no
the origami paper changed between the videos	yes	no

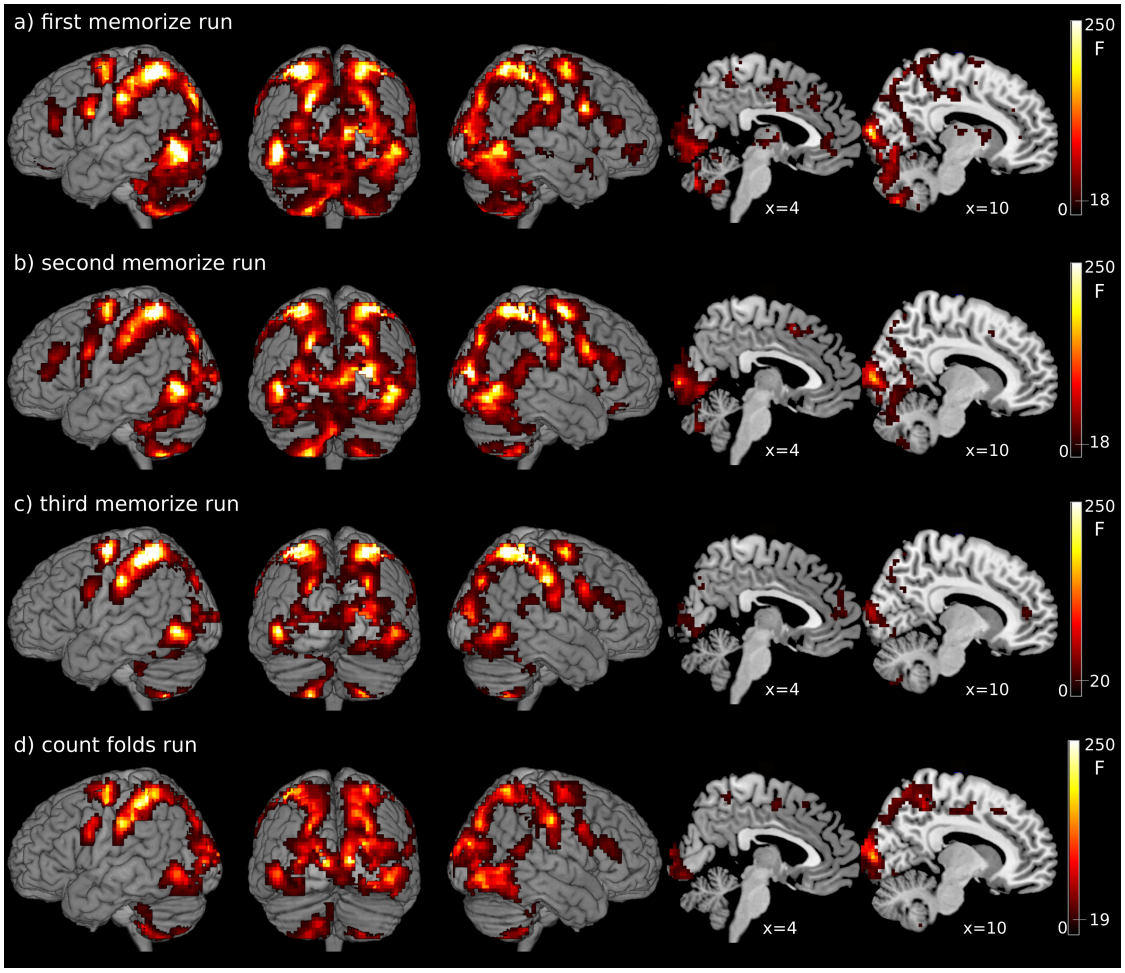


Figure 2.10: Significant coupling between instructor and observers for each condition. Thresholded at $p<0.001$, $p_{FWE,cluster}<0.05$. This threshold is also fulfills $q<0.01$ FDR corrected at the voxel level.

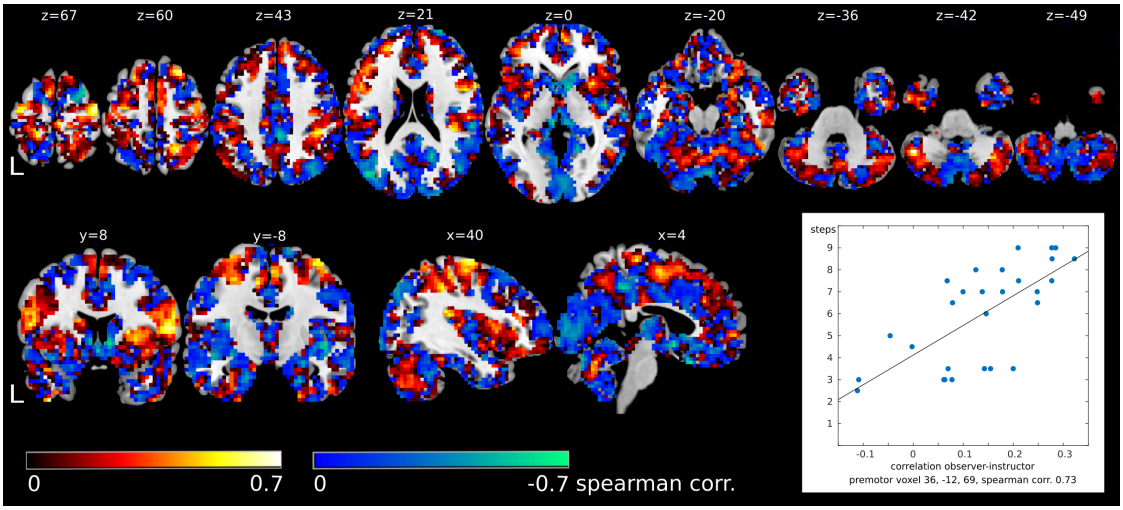


Figure 2.11: Whole-brain-analysis of the correlation between instructor-observer synchronization and subsequent observer performance for the first learning run. No cluster of significant size as utilized in this paper survives FDR correction or a display at uncorrected p<0.001.

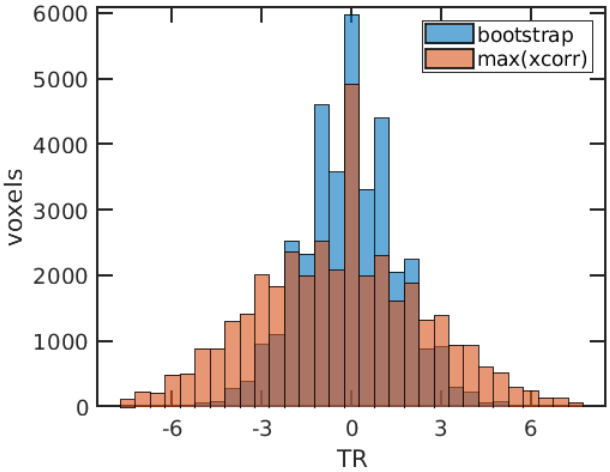


Figure 2.12: Comparison between the median delays for the subsampling bootstrap routine and a classical routine searching for the maximum of the cross-correlation within the complete time course. Blue bars analog to fig. 2.8. Note that the estimated median lags are more spread out to more extreme delays for the maximum cross-correlation routine.

3

Synchrony between observers during observation and imitation learning of origami

Kathrin Kostorz, Virginia L. Flanagin, Stefan Glasauer

Abstract Online video learning is common practice nowadays. In this study, using fMRI we investigated the temporal synchrony between subjects during observation of instructive origami videos with either the task to learn the steps or, as a control, to count folds. We assessed the similarity between the observers using principal component analysis. High similarity, as described by a high explained variance of the first component, was found in typical areas of the action observation network and cerebellar hand motion areas for all different tasks of learning or identifying actions. Within subjects we could show similar processing of the same video as assessed by intrasubject correlation for all different learning conditions in visual areas and the AOEN, across different videos in visual and parietal areas. Changes in synchrony between sessions were found especially in higher visual areas, the TPJ, parts of the premotor cortex, and large parts of the cerebellum. Last but not least, we determined the optic flow of the videos. A correlation of the optic flow with the first principal component revealed a very high agreement within the AOEN, in particular the somatosensory cortex, which possibly reflects hand motion processing in these areas. Finally, our data indicates the involvement of the primary motor cortex in action observation, with a low but significant and consistent contribution. Taken together, we present new aspects of the AOEN during learning and action processing; in particular our data stresses the involvement of the cerebellum in diverse tasks, well beyond motor control areas.

3.1 Introduction

Nowadays, when someone is interested in 'how something is done', this information can be found easily and quickly by searching online video platforms for it. This way one can observe an instructor performing the desired actions or processes, in order to understand the process and to imitate it if wished. Observing online instructive videos has thus become a common way of learning. This common practice is also more effective than previous methods: Michas and Berry (2000) and Wong et al. (2009) found out that video or animated instructions are superior to text or static instructions; Marcus et al. (2013) reported that with-hands animations are superior to without-hands animations. Consequently, learning by observing online videotaped instructions is an effective way of learning procedures or skills.

Classical action observation in the human brain is associated with the so-called action observation network (e.g. Grafton, 2009) or action observation execution network (AOEN) (Thomas et al., 2018), which is also involved in the observational learning of actions (e.g. Lago-Rodríguez et al., 2014). It consists of visual (V5), parietal, somatosensory, premotor and cerebellar areas. Using functional magnetic resonance imaging, the AOEN has been shown to modulate activity with the training and expertise of the observer (e.g. Calvo-Merino et al., 2005, 2006) during action observation. In particular, a higher AOEN activity for one's own intensively trained actions (as in dancers) has been reported. For sequence learning as a specific form of skill learning (not necessarily by observation), activity increases as well as decreases that are associated with several days of training have been reported (e.g. Steele and Penhune, 2010; Penhune and Steele, 2012; Wiestler and Diedrichsen, 2013). These could be associated with increased neural recruitment for trained behaviors as well as reduced activity due to higher neural effiency (Wiestler and Diedrichsen, 2013). Nevertheless, even in reduced activity, sequence specificity has been shown (Wiestler and Diedrichsen, 2013). Importantly, these studies utilize training blocks of several days. Classically, mastering a skill needs a lot of practice and refining as skill-learning can be characterized into an initial, fast learning stage, followed by a longer-lasting, but slower learning stage (Dayan and Cohen, 2011).
For the initial stage of observational skill-learning, i.e. the initial observation of actions and immediate attempts to reproduce them, the imaging literature is surprisingly sparse (Sakreida et al., 2017). This is partly due to the fact that movement options are very limited inside the MRI machine, also leading to potential motion artifacts in the recordings. Buccino et al. (2004); Vogt et al. (2007); Higuchi et al. (2012) have investigated the observation and immediate reproduction of guitar chords, Sakreida et al. (2017) of keyboard presses. They have shown a higher involvement of the AOEN during imitation than observation, decreases of AOEN activity after several days of practice, and postulated the dorsolateral prefrontal cortex (DLPFC) as a cognitive control hub for

observational imitation learning. For observational learning of a longer action sequence, Frey and Gerry (2006) have investigated the immediate reproduction of a toy model and showed a higher AOEN activity for reproduction of the exact sequence than for outcome reproduction. Consequently, while we know that the AOEN is involved in the initial observational learning, it is unclear how this involvement changes with the initial trials. Will we get already lower involvement as after several days of practice or higher involvement due to better action understanding? Also, we know that the AOEN is more greatly involved during observation for requirement of reproduction than pure observation (Buccino et al., 2004; Frey and Gerry, 2006), but pure observation does not require the active processing of the stimulus. Consequently, it will be interesting to compare observation with the requirement of reproduction and observation with another action identification task.

Investigating the initial stage of naturalistic observational learning requires a paradigm beyond the classical MRI paradigms. Normally, these videos show instructive sequences of actions that one watches and tries to imitate afterwards. Consequently, it will be important to develop a setup where it is possible to watch the instructive videos inside the MRI machine and reproduce them straight away, both tasks alternating. Moreover, one needs new data-driven analysis methods which, as we show in this paper, are helpful in analyzing the longer (instructional) videos. Additionally, adapting data-driven approaches of video content and features and relating them to the neural activity will be informative: It will give us information about the neural processing of these contents of the video in a naturalistic, everyday environment.

Free viewing of feature movies is by now an established fMRI paradigm (e.g. Hasson et al., 2004, 2008a; Hasson and Honey, 2012). Typically, similarities and commonalities of the subjects' BOLD signals are searched for here. The rationale for these open experiment designs is the following: if the measured BOLD signal is time-locked and similar between subjects in corresponding voxels or areas we assume a common processing of the stimuli and therefore common cognitive processes. A classical measure of this similarity is the intersubject correlation method (ISC) introduced to fMRI by Hasson et al. (2004). Typically, studies employing the ISC method engage feature movies and parts thereof, and contrast ratings of features (e.g. Nummenmaa et al., 2012) or scrambled parts (e.g. Hasson et al., 2008b). However, a typical how-to video offers reduced visuals in comparison to feature movies: it is often not or only slightly edited and typically involves only one camera perspective, more like every-day life visual stimuli. This has consequences: Herbec et al. (2015) show that edited videos lead to higher ISC than non-edited ones. Consequently, it is especially important to also investigate unedited videos if one wants to make inferences about everyday life.

Using ISC or a familiar method, intersubject phase synchronization (Nummenmaa et al.,

2014), two papers have investigated the action observation and observer synchrony. Nummenmaa et al. (2014) have shown a synchronization of observer BOLD activity in parts of the AOEN (parietal, somatosensory and premotor cortex) during observation of boxing videos when the subjects mentally simulate a boxer. Moreover, Thomas et al. (2018) have shown the involvement of the AOEN during movies with sequences of everyday actions using ISC. They have shown that a meaningful order leads to higher involvement of parts of the AOEN (in particular parietal areas). It is interesting to investigate how this synchrony in the AOEN translates to the observation of how-to videos and observational learning. Additionally, extracting the visual motion of the videos and relating it to the neural activity of the AOEN will show new aspects of specific action processing of uncontrolled naturalistic hand movements during observational learning.

To our knowledge, no one has investigated observational learning when watching a how-to video, looking for similarities between subjects. By requiring immediate reproduction we want to keep a naturalistic learning setup. As movement options are limited inside the MRI machine and potential head motion can corrupt the data severely, we have chosen the creation of a small object: by observing instructional foldings of paper the subjects should learn to reproduce a paper figure, also known as an origami.
The aim of this study was threefold: first to investigate similarities between observers when observing how-to videos. We will show synchrony changes in relation to the subsequent first steps of observational learning and a comparison of synchrony between observing a video with the intention to reproduce in contrast with another action identification task which requires attention to the video: to count how many folds are made. The second aim of the study is to establish a different approach to the temporal similarity calculation: The principal component analysis. This approach is very similar to the ISC analysis but investigates the commonness of all subjects at once, without forming pairs. Moreover, the different characteristics of the PCA can be utilized to characterize different aspects of the commonness beyond pure synchrony as we will show. Finally, we present the visual optic flow of the videos as a commonality in processing and as a putative proxy for the shown uncontrolled hand-movements of how-to videos.

3.2 Materials and methods

3.2.1 Experimental setup and procedure

The general setup of the experiment has been described in chapter 2. Participants, stimuli, MRI setup, experimental procedure, fMRI acquisition and the fMRI data preprocessing are identical to this description. Therefore, we just give a brief illustration of the setup here. For technical details, more comprehensive information about the experimental procedure as well as detailed information about the instructor, please refer to chapter 2 section 2.2.

Subjects

For this study we recruited thirty-three right-handed subjects to become the observers. We required that the subjects had no experience with complex origami figures within the last 5 years. Here, complex means anything other than traditional paper figures like a plane or boat. After dropouts, 28 subjects were analyzed (14 female, mean age 27.2, range 21 to 36 years).

Stimuli

Two different paper-folding videos were shown during the experiment. The origami to be learned was a butterfly consisting of 8 or 9 folds (depending on whether an additional optiona stabilizing fold was made). Importantly, the final shape of the butterfly could only be recognized during the last fold. This way, we wanted to make sure that the subjects could not complete the figure by just memorizing the final shape. As a control condition we showed a video of partly the same or similar folds, but partly repetitive. This way, the folding does not lead to a final origami, but ends with an open sheet of paper. We will refer to this control condition as 'unproductive folding'. Both videos were performed by the same instructor (see chapter 2) in an instructive way. This means the video was moderately paced and after each folding the corresponding intermediate step was clearly shown, as it is important that the hands do not hide the paper. The length of the shown butterfly video was 03:25 min, that of the control condition was 03:32 min. The videos were monochrome.

MRI setup

The whole experiment was performed inside the MRI machine. The subjects were shown the videos on a screen positioned behind the MRI machine bore, visually accessible through one mirror of the standard Siemens MRI mirror system (the videos were digitally mirrored before demonstration to account for the laterality flip). On the subject's pelvis

we positioned a cardboard trapezoidal folding table as well as a stack of gray origami paper (size 15x15 cm). As head movements are crucial during MRI experiments, arms were raised by foam cushions so that ideally, subjects could perform the folding only moving their hands from the wrists onwards. Through the other double mirror of the Siemens mirror system (MS) the subjects could observe their own hand movements on the folding table. The experiment alternated between watching the video and folding the steps which were memorized. As the MS only allows for one perspective (either screen or hands), the subjects were instructed to shift the MS on the rails of the headcoil in between the conditions. They were instructed to do so slowly and without moving the head and reminded of this before every shift. On the MS we mounted a video camera which recorded the performance of the subjects during the folding attempts. The instructive videos were taken with the same setup; consequently the perspective of the subjects when folding is almost the same as in the videos.

Experimental procedure

Figure 2.1 (chapter 2 section 2.2) shows the experimental procedure of the experiment. When regarding the figure, please note that in this manuscript, only the observers are analyzed. For the relation between observers and the instructor, please refer to chapter 2.

Before the main experiment started, the subjects were shown the final origami they would have to fold. This way we wanted to ameliorate a potential surprise effect. After being brought into the final position on the MRI table, we performed a short familiarization with the subjects: the participants were instructed to fold an exemplary sheet of paper (in any way they wished) to get used to the setup. Importantly, this way they were able to familiarize themselves with seeing their hands through the mirror system.

During the main experiment, the subjects had three attempts to watch the butterfly video with the task to memorize the steps and reproduce afterwards. During the control condition, the subjects watched the video of 'unproductive folding' with the task of counting the number of folds being made. Right afterwards they had to report the number via the MRI audio communication system. This was to assure the attentiveness of the subjects to the video as no reproduction was required here. The control condition was randomized with the whole origami-learning block.

During the learning runs, subject had to watch the instructive videos with the task of 'memorizing as far as you get'. Subjects were instructed to not move their hands during this. Right after the video ended, they had to reproduce the foldings 'as far as they get' and 'indicate with their hands that they are done' (i.e. make a waving sign). After each folding session, they had to watch the video again, and then fold again, resulting in three observe-fold combined runs.

fMRI acquisition and preprocessing

Scanning was performed on a Siemens Skyra 3T MRI machine using a 64-channel head coil. Before the beginning of the main experiment, a field map was recorded. For the EPI sessions, a voxel size of 2.5x2.5x2.5 mm and a repetition time of 0.59 seconds was achieved. The first 11 scans were discarded; during this time, a black screen with a white fixation cross was shown (with no task given; subjects were informed about the purpose of this.) In the end, the observation run of the butterfly origami had 349 images, that of the control condition had 362. After the main experiment, a structural image was taken for the later coregistration and normalization routine.
Preprocessing of the data was performed using SPM12 and FSL 5.09 for the high-pass filtering. The preprocessing included calculating a voxel displacement map, realigning and unwarping the functional data, coregistration of the anatomical image, segmenting the anatomical image and creating a common template using DARTEL (Ashburner, 2007) to ensure the best fit. The functional data was normalized to 3x3x3 mm using a gaussian smoothing kernel of 6x6x6 mm full-width at half maximum (FWHM). FSL was then used for highpass filtering of the functional data with a cutoff of 100ms.

3.2.2 Data analysis

Principal component analysis

To investigate the intersubject similarity and synchrony, we expand the classical intersubject correlation (ISC) (e.g. Hasson et al., 2004) approach by performing a principal component analysis (PCA) on the data set (see e.g. Abdi and Williams 2010b for a detailed description of this dimensionality reduction technique). Essentially, the principal components describe the commonness between the different data along orthogonal directions - in our case the commonness between all subjects' time courses per voxel. Typically, the principal components are sorted in descending order according to how much variance of the data is explained by the component (i.e. the first component is the one explaining the most variance of the data set). If there is a high synchrony between the subjects, the first component will explain much of the data variance; essentially, this is a measurement of 'how much everybody digs into the same direction' (Friedman and Weisberg, 1981). As our simulations have shown, for positive correlations the explained variance of the first principal component shows a good approximate linear relationship to the average pairwise correlation coefficient, which is in line with the findings of Friedman and Weisberg (1981). This way, we can ensure sure that previous results remain comparable to this method. Figure 3.8 shows the results of the simulations.
For the intersubject correlation analyses, there are typically two approaches: Either cal-

culate the pairwise pearson correlation between all subjects' time courses, then average the correlation over pairs (e.g. Kauppi et al., 2014). Or compute per subject the mean time course of all other subjects, and then correlate the respective two time courses (subject and mean), see e.g. Hasson et al. (2008b). In contrast to the classical intersubject correlation approach, for the PCA approach no computing of pairwise correlations between subject pairs or averaging is needed (for further information about these arising problems of the classical ISC approach see Chen et al., 2016a). A precursory approach to ours has been described in the context of brain synchrony measured by fMRI by Hanson et al. (2009). They performed a PCA on the pairwise correlation matrix and describe the clustering of subjects data near the principle components of that matrix, especially the first one. This offers a measure for temporal synchrony which is the largest eigenvalue of the correlation matrix of the data set. In this paper we generalize the approach, as mathematically no correlation matrix needs to be constructed, and also utilize additional metrics the PCA offers.

The PCA approach offers three characteristics to describe the data: common time courses of the subjects along the components (in Matlab referred to as 'scores'), the strength of each component (the variance explained by each component) and the loadings (in Matlab referred to as 'coefficients', sometimes also referred to as 'latent variables'). The displayed 'explained variance' in this manuscript is the normalized explained variance: the explained variance of the component divided by the total variance (sometimes also referred to as separability index). The loadings/coefficients describe the contribution of each subject to the component.

To compute the PCA we use a singular value decomposition (SVD) of the whole data set. All analyses were performed in Matlab version R2018a using the in-built PCA function (with SVD as the algorithm of choice). Every statistical significance in this paper is assessed via bootstrap tests (see the following paragraphs for details).

A note of caution should be addressed when using the PCA framework: by definition, the direction of the principal component (meaning the overall algebraic sign) is per se arbitrary for the results of the algorithm. In Matlab, the coefficients (sometimes also called 'latent variables') are chosen in such a way, that the overall contribution remains positive (meaning the sum of the coefficients is positive). In case of weak components, this could mean that few 'strong' subjects determine the algebraic sign of the component and therefore also that of the resulting score.

Data synchrony for each session To assess whether there is a common similarity between the observing subjects for all conditions, including the control of counting folds, we computed the PCA for each session separately. Since the components are ordered by descending explained variance, there is a common similarity for the data set if the first component (i.e. the biggest component) is significant. Assessment of significance of

the components was done via a bootstrap under the null hypothesis that the temporal matching of the subjects' time courses is irrelevant: the null data was generated by circularly shifting each subject's data by a random number. This keeps the temporal data structure overall (and therefore the autocorrelation of the signal), but violates the temporal alignment between subjects. It is an established routine for the analysis of intersubject correlation, as also employed in the ISC toolbox of Kauppi et al. (2014). On the shifted data, we performed the PCA to obtain a null distribution of explained variances for significance testing. We performed 10000 iterations of the bootstrap. This procedure can be used to test all components; in practice for our data set almost always only the first component offered a significant explained variance.
The data were thresholded at $q<0.001$ FDR corrected and for an (arbitrary) extent threshold of 19 voxels. For this thesis, we used a slightly flexible extent threshold of 17-20 voxels as no cluster correction can be applied on the not-smooth-enough bootstrapped p-values. This way, all relevant information but no too-small clusters were displayed.
To show the voxelwise synchrony present over all conditions, we calculated a conjunction of all images by creating a common minimal mask.

Differences in data synchrony We also wanted to determine the changes in synchrony between the observers between the counting and learning condition as well as between the first and last learning condition. To incorporate the repeated measures design into the statistics, for the purpose of this thesis we calculated the differences in synchrony via the established method used in intersubject correlation studies (see e.g. Lerner et al., 2011; Regev et al., 2013 or Thomas et al., 2018). We calculated the correlation between one subject's time course and the average of the time courses of all other subjects for the corresponding session. The test-statistic then is the difference between the z-transformed correlations for each subject, analog to the paired t-test. The real test-statistic distribution was estimated via a bootstrap routine: random subsamples of the test values were drawn with replacement for this. As the null hypothesis was 'no difference', p-values were determined by the number of incidents smaller (or for negative median bigger) than '0'. The test was performed as a two-sided test with 10000 iterations.

Inter-session correlation/intra-subject correlation

Apart from changes in synchrony between sessions, we also wanted to assess the similarity between the sessions. Therefore we correlated each subject's time course for one session with the same subject's time course for the other session. This has been described as intra-subject correlation Golland et al. (2007). Note that we did not perform a PCA as this is a within-subject statistic and does not ask 'how much does anybody dig into the same direction as everybody else in one condition'. Also, only n pairs of correlations

are computed, with every measurement being independent of the other. For statistical testing, we shifted the time courses of each subject circularly to the other condition of that subject, destroying the time-locking. Note that also in this case, this or a similar kind of nonparametric testing is necessary, as the tests are performed within subjects; therefore, correlations due to similar physiology are to be expected. The circular shifting should also conserve the temporal structure of these confounds. The null hypothesis of this test states that the temporal relation between the two sessions of one subject is irrelevant to the correlation between the sessions.

Optic flow

We computed the optic flow of the videos using the Horn-Schunk algorithm using MATLAB. The results were then convolved with the standard hemodynamic response function of SPM and afterwards downsampled to the TR of the MRI images. To assess processing of the optic flow in the subjects, we correlated the parameter with the timecourse of the first principal component/score of each voxel. For statistical testing, we circularly shifted each subject's time course, recomputed the first score and it's correlation with the optic flow. As the direction of the score is per se arbitrary (see explanation above) we also checked that for the displayed significant correlations, not only the main contribution to the component is positive but also that the majority of subjects contribute to the component in a positive way. This way we can make sure that for the negative correlations with the optic flow the sign of the resulting common description of brain activity is supported by the majority of the subjects.

3.3 Results

3.3.1 Overall similarity during action observation

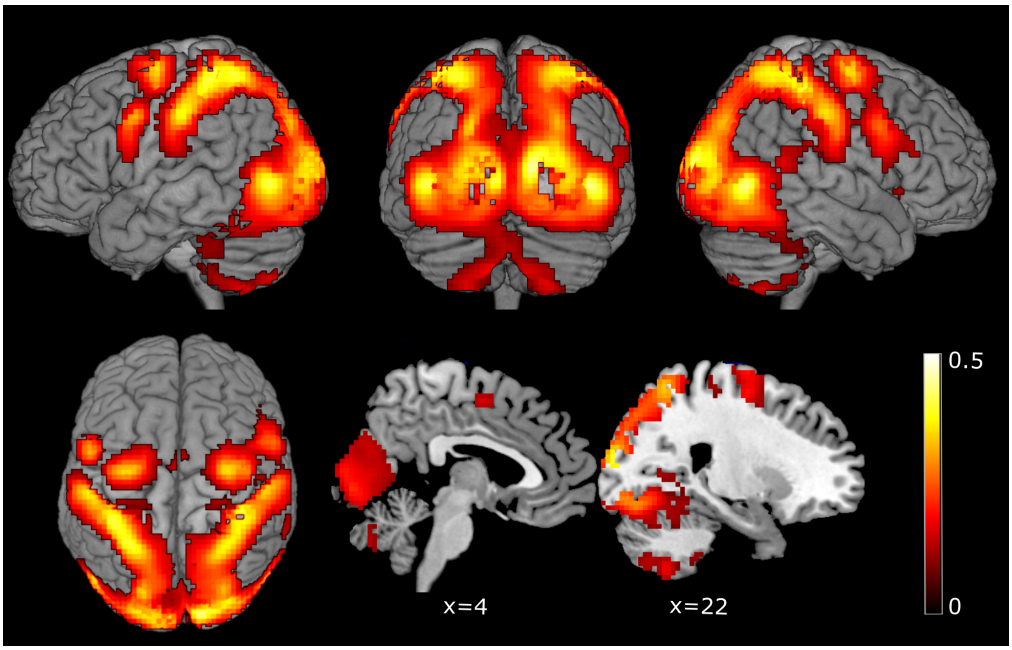


Figure 3.1: Voxels where the variance explained by the first component is significantly different from '0'. The AOEN is significantly synchronized between the observers for all conditions of this experiment. For display reasons, average explained variance over all conditions is displayed. Conjunction done on single condition images, each q<0.001 FDR thresholded. Only clusters $\geq$ 19 voxels displayed here.

First, we investigated the voxels where the explained variance of the first component is significantly different from the variance explained by the null condition of temporally non-matched time courses (see materials and methods). For this analysis we performed a separate PCA on each observation run, i.e. the three memorize runs and the count folds run, and determined the significance of the first component as a measure of synchrony. Figure 3.1 shows a conjunction of all significant voxels of all conditions. This means that these voxels showed a common significant explained variance for all three learning runs as well as the control run of counting folds. We find a robust synchrony between the observers over all conditions in classical AOEN areas. This means that the observer's neural signals are synchronized during all three learning runs as well as the count folds run in the following areas: A collective principal component was significantly present in the inferior parietal lobe (IPL), superior parietal lobe (SPL), primary somatosensory voxels, dorsal and ventral premotor voxels and the supplementary motor area (SMA). These are classical areas of the AOEN. Additionally, a collective component was present in typical visual/occipital areas and the right TPJ. Moreover, cerebellar lobule VI, VIIa and VIIb,

location	cluster size	peak statistics x	y	z	param.	peak location *area with max probability*
conjunction of all significant first components, q<0.001, k≥20						
R+ L Cerebellum (VI, VIIa+b, VIIIa+b), Occipital Cortex, SPL, IPL, Area 44, Area 4a, Area 4p	9547	12	-93	9	0.45	R Calcarine Gyrus; Area hOc1 [V1] 37%
		45	-66	0	0.44	R Middle Temporal Gyrus; Area hOc5 [V5/MT] 49%
		-9	-96	3	0.42	L Superior Occipital Gyrus; Area hOc1 [V1] 58%
		39	-33	54	0.42	R Postcentral Gyrus; Area 3b 51%
		-45	-72	0	0.41	L Middle Occipital Gyrus; Area hOc5 [V5/MT] 54%
		-39	-39	60	0.41	L Postcentral Gyrus; Area 2 42%
		-36	-42	63	0.40	L Postcentral Gyrus; Area 2 40%
		-30	-51	63	0.40	L Superior Parietal Lobule; Area 7PC (SPL) 51%
		39	-39	60	0.38	R Postcentral Gyrus; Area 2 38%
		-51	-24	42	0.38	L Supra Marginal Gyrus; Area 2 56%
		30	-51	63	0.38	R Superior Parietal Lobule; Area 7PC (SPL) 66%
L Precentral Gyrus	348	-33	-9	66	0.35	L Precentral Gyrus
L Precentral Gyrus, Area 44	217	-54	3	39	0.31	L Precentral Gyrus
R SMA	37	3	-3	54	0.15	R Posterior-Medial Frontal
L SMA	32	-3	-3	54	0.16	L Posterior-Medial Frontal
R Insula	26	33	24	-3	0.12	R Insula Lobe
R Precentral Gyrus	22	24	-27	63	0.12	R Precentral Gyrus

Table 3.1: Conjunction of all PCAs for all runs, corresponding to fig. 3.1. 'param.' denotes the displayed mean variance explained. Note that the conjunction was performed as a 'physical' conjunction, with each condition fulfilling a threshold of q<0.001 FDR corrected.

and VIIIa and VIIIb showed common activations. Finally, we report significant explained variance in areas 4a and 4p, which are typically associated with primary motor function. Note that the explained variance was low in these areas compared to the classical AOEN areas.

Figures 3.6 and 3.7 show the significant explained variance for each condition separately. Additionally to the commonness for all conditions, we found significant components in the superior temporal sulcus (STS, first memorize condition), BA 45 and DLPFC (all memorize runs), thalamus (first memorize run), basal ganglia (first memorize run), posterior and middle medial frontal cortex (first and second memorize, count folds run) and medial prefrontal cortex (second and third count folds run).

3.3.2 Commonness between sessions

We wanted to examine the commonness in processing during the different video runs, also given the varying cognitive tasks (learning, monitoring, counting). For this we determined the intrasubject correlation between the first and third memorization run, as well as between the first memorization run and the count folds run. In the two memorize runs, the same video was shown, but in the first run, it was new and performance variable, for the third run most people have managed to fold the origami. The videos shown were different for the comparison between the first memorize run and the count folds run. Important for this comparison is the fact that both have been viewed for the first time

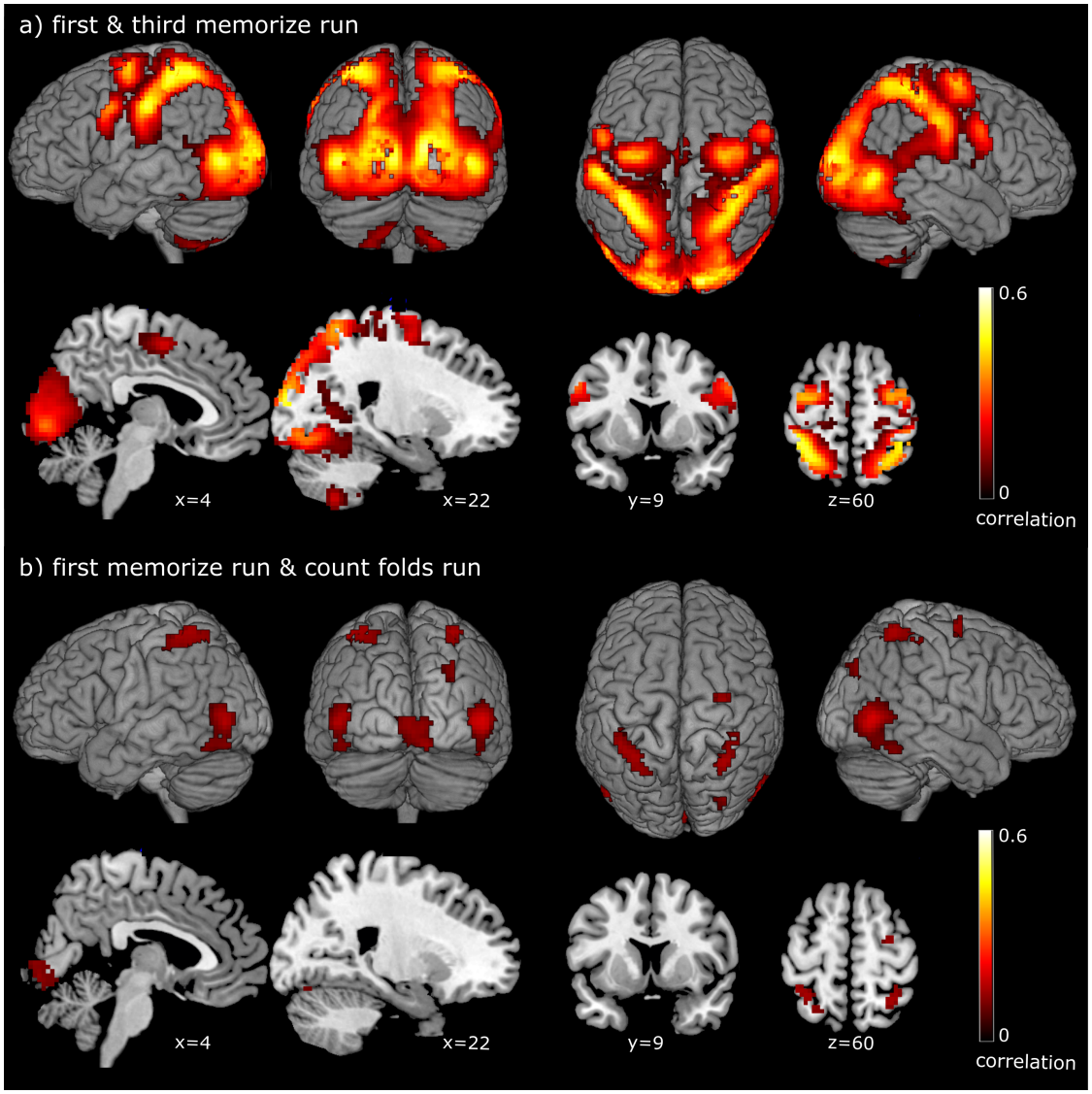


Figure 3.2: Significant intrasubject correlations. High similarity in AOEN within subjects between the memorizing runs, low similarity between memorizing and control run. a) intrasubject correlations between the first and the third memorization run. $q<0.001$ FDR corrected, $k\geq19$. b) Intrasubject correlations between the first learning run and the count folds run. $q<0.01$ FDR corrected, $k\geq19$. Note that at $q<0.001$ nothing is significant at our number of bootstrap iterations. But note also that in both cases a) and b), the resulting uncorrected probability threshold is $p=0.0002$.

and were counterbalanced.

When investigating intrasubject correlation between the first and third memorization run, we found a high level of similarity in the occipital cortex and the classical AOEN areas SPL, IPL, dorsal and ventral premotor cortex as well as somatosensory cortex and area 4; these areas also showed relatively high correlations of around 0.4-0.5 (see fig.

location	cluster size	peak statistics x	y	z	stat.	peak location *area with max probability*
intrasubject correlation first and third memorize run						
R+L Occipital Cortex,	9228	15	-93	12	0.53	R Cuneus; Area hOc2 [V2] 35%
SPL, IPL,		-9	-96	3	0.51	L Superior Occipital Gyrus; Area hOc1 [V1] 58%
Postcentral Gyrus, Area 4		-45	-72	3	0.50	L Middle Occipital Gyrus; Area hOc5 [V5/MT] 52%
		-39	-39	60	0.50	L Postcentral Gyrus; Area 2 42%
		-54	-21	39	0.49	L Inferior Parietal Lobule; Area 1 42%
		45	-66	0	0.49	R Middle Temporal Gyrus; Area hOc5 [V5/MT] 49 %
		-36	-42	63	0.49	L Postcentral Gyrus; Area 2 40%
		21	-93	15	0.49	R Superior Occipital Gyrus; Area hOc3d [V3d] 35 %
		-30	-51	63	0.49	L Superior Parietal Lobule; Area 7PC (SPL) 51 %
		-48	-24	39	0.48	L Inferior Parietal Lobule; Area 2 61%
		39	-33	51	0.47	R Postcentral Gyrus; Area 2 48%
R Cerebellum	158	21	-60	-54	0.18	R Cerebellum (VIII); Lobule VIIIb (Hem) 69 %
		12	-72	-48	0.14	R Cerebellum (VIII); Lobule VIIIa (Hem) 74 %
		27	-42	-51	0.11	R Cerebellum (VIII); Lobule VIIIb (Hem) 71 %
L Cerebellum	154	-9	-72	-48	0.15	L Cerebellum (VIII); Lobule VIIIa (Hem) 78 %
		-24	-54	-54	0.14	L Cerebellum (VIII); Lobule VIIIb (Hem) 60 %
		-18	-69	-51	0.14	L Cerebellum (VIII); Lobule VIIIa (Hem) 84 %
SMA	133	3	-3	54	0.18	R Posterior-Medial Frontal
		-3	-3	54	0.17	L Posterior-Medial Frontal
intrasubject correlation first memorize run and count folds run						
R Occipital Cortex (V5)	248	51	-66	3	0.18	R Middle Temporal Gyrus;Area hOc5 [V5/MT] 63 %
		48	-51	-21	0.11	R Inferior Temporal Gyrus; Area FG4 52 %
		48	-63	-15	0.1	R Inferior Occipital Gyrus; Area FG2 53 %
L Occipital Cortex (V5)	174	-48	-72	0	0.15	L Middle Occipital Gyrus; Area hOc4la 59 %
		-48	-72	-15	0.11	L Inferior Occipital Gyrus; Area FG2 57 %
		-48	-63	-15	0.11	L Inferior Occipital Gyrus; Area FG2 57 %
R+L Medial Occipital Cortex	152	12	-84	-15	0.13	R Cerebellum (VI); Area hOc2 [V2] 50 %
		3	-87	-6	0.13	L Calcarine Gyrus; Area hOc1 [V1] 66 %
		3	-87	-6	0.13	L Calcarine Gyrus; Area hOc1 [V1] 84 %
		9	-87	-3	0.12	R Lingual Gyrus; Area hOc1 [V1] 64 %
L SPL, Area 2	117	-33	-48	63	0.14	L Superior Parietal Lobule; Area 7PC (SPL) 72 %
		-39	-45	63	0.14	L Postcentral Gyrus; Area 2 35%
		-42	-39	60	0.14	L Postcentral Gyrus; Area 3b 21 %
R SPL, Area 2	83	30	-51	63	0.15	R Superior Parietal Lobule; Area 7PC (SPL) 66 %
		36	-39	57	0.11	R Postcentral Gyrus; Area 2 78 %
R Superior Frontal Gyrus	20	30	-9	66	0.13	R Superior Frontal Gyrus
R Superior Occipital Gyrus	19	30	-78	39	0.11	R Superior Occipital Gyrus

Table 3.2: Significant intrasubject correlations for the first and third memorize run; and for the first memorize run and the count folds run. Corresponding table to figure 3.2. 'stat.' denotes Fisher z-values. q<0.001 FDR corrected.

3.2 and table 3.2). Also cerebellar areas VIIIa and VIIIb showed significant intrasession similarity. Consequently in these areas each subject showed high self-similarity during action observation irrespective of subsequent performance.

For the intrasubject correlation between the first memorize run and the count folds run, we found significant correlations in primary visual areas. Moreover, considering the AOEN, area V5, the SPL and parts of area 2 showed significant intrasubject correlations. These correlations were low, ranging between 0.1 and 0.2 (see table 3.2). Therefore each subject displayed low self-similarity during observation to reproduce and observation to count folds in the aforementioned areas.

3.3.3 Difference between sessions

Apart from the commonness between runs, we also wanted to investigate the changes in synchrony between them. Overall, when comparing the first memorize run to the third memorize run or the (counterbalanced) counting folds run, we mainly find a reduction in synchrony. That means that between-observers synchrony is overall higher in the first memorize run.

Synchrony changes between the first memorize run and the third memorize run We compared the first memorize run with the third one to compare pure memorization to (mainly) monitoring of the already known steps, see fig. 3.3a) and table 3.3. The biggest changes in correlation of one subject to his/her fellow participants between these sessions appeared in visual areas as well as areas of the dorsal stream, in the bilateral posterior middle temporal gyrus reaching to the temporo-parietal junction (TPJ), in the superior temporal sulcus (STS) and in medial superior parietal areas/parts of the precuneus. For parts of the classical AOEN, we found the highest changes in lateral parts of the premotor cortex. We also found changes in the DLPFC and IFG (BA45). Last but not least, great extents of the cerebellum showed significant reduction in synchrony, especially ones not classically related to motor tasks like the crus1 and crus2.
While the extent of areas showing synchrony decrease from first to third run is vast, some areas showed higher synchrony for the third run: in this case medial prefrontal areas demonstrated higher synchrony between the subjects.

Synchrony changes between the first memorize run and the count folds run When comparing the first memorize run to the control condition of counting folds, partly similar changes to the contrast of the first and third run can be reported, see fig. 3.3b) and table 3.4. In particular, we found higher synchrony for the memorize condition in visual and dorsal stream areas, bilateral posterior middle temporal gyrus to TPJ and supramarginal gyrus, medial superior partietal lobe and STS. Of the classical AOEN areas, we can report higher synchrony for memorizing in two bilateral premotor clusters and BA 44. Also for this comparison, the cerebellum showed higher similarity between subjects to a great extent in the memorize condition.
On the other hand, for counting folds the subjects showed higher synchrony in a small cluster of the visual cortex.

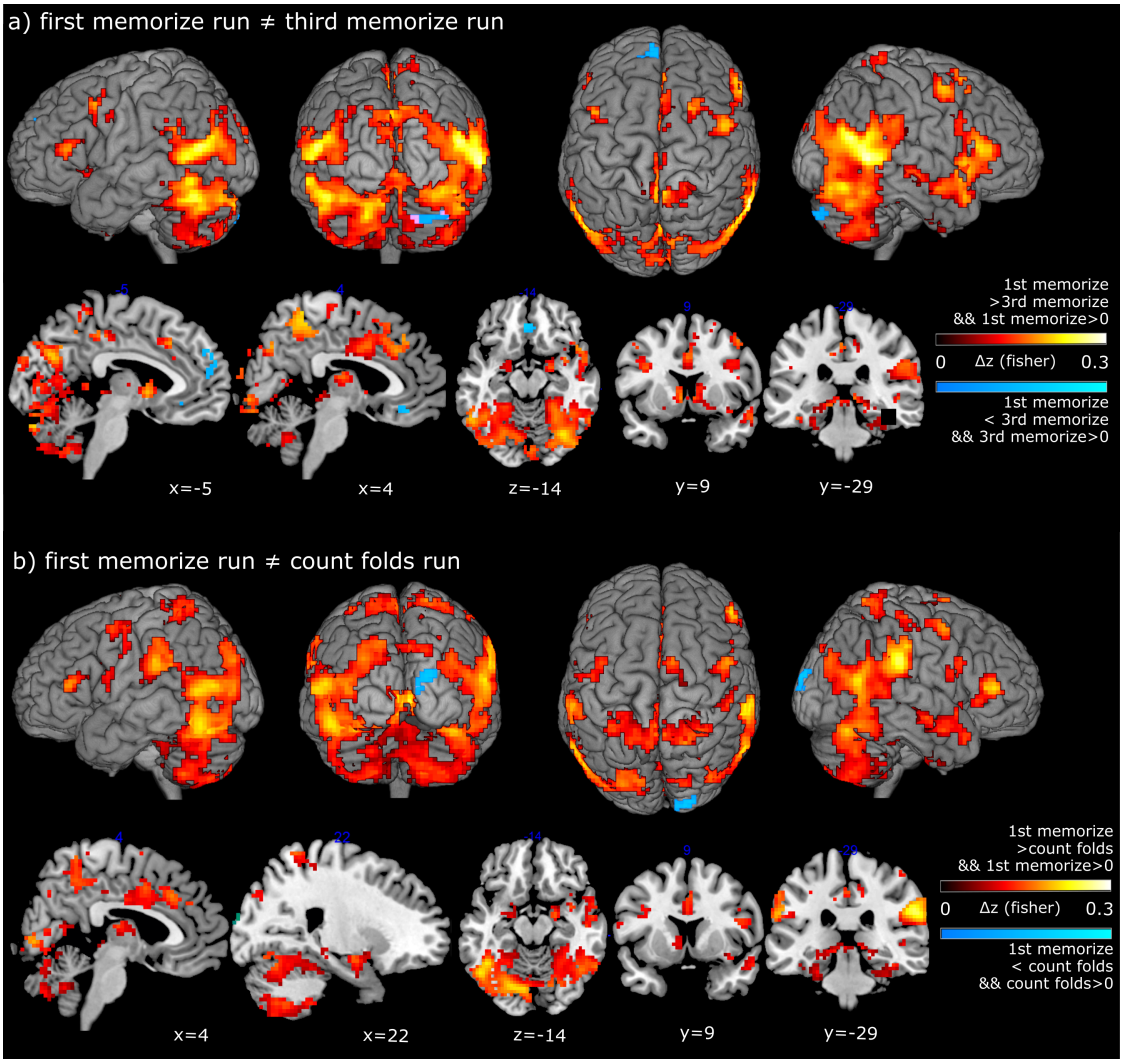


Figure 3.3: Significant synchrony changes. Overall higher between-observer synchrony in the first learning run than the third learning run and the count folds run. a) Synchrony changes between the first and the third memorization run. $q<0.001$ FDR corrected, $k\geq17$ ($k=17$ is the cerebellar cluster, otherwise $k\geq19$. b) Synchrony between the first learning run and the count folds run. $q<0.001$ FDR corrected, $k\geq19$. Cluster threshold arbitrary.

location	cluster size	peak statistics x	y	z	stat.	peak location *area with max probability*
first memorize>third memorize ∩ first memorize>0						
R+L Cerebellum (V,VI,Crus1+2,VIII), Occipital Cortex, SPL, IPL, R BA44,45, R Amygdala	5925	63	-51	3	0.29	R Middle Temporal Gyrus
		48	-63	18	0.27	R Middle Temporal Gyrus
		57	-57	12	0.26	R Middle Temporal Gyrus; Area Pga (IPL) 45%
		-48	-60	-21	0.25	L Cerebellum (Crus 1); Area FG2 51%
		33	-69	-18	0.23	R Fusiform Gyrus; Area FG1 47%
		-60	-57	-9	0.23	L Inferior Temporal Gyrus
		-18	-78	-36	0.22	L Cerebellum (Crus 2); Lobule VIIa crusII (Hem) 79%
		48	-48	21	0.22	R Superior Temporal Gyrus
		57	30	9	0.21	R IFG (p. Triangularis); Area 45 70%
R+L SPL (Area 5)	617	3	-45	63	0.21	R Precuneus; Area 5M (SPL) 56%
		3	-42	54	0.21	R Precuneus; Area 5M (SPL) 66%
		-3	-45	60	0.20	L Precuneus; Area 5M (SPL) 69%
		-3	24	39	0.19	L Superior Medial Gyrus
		3	36	42	0.18	L Superior Medial Gyrus
		-3	-33	48	0.18	LMCC; Area 5M (SPL) 41%
L IPL MST, TPJ	486	-54	-63	15	0.26	L Middle Temporal Gyrus
		-57	-54	6	0.24	L Middle Temporal Gyrus
		-45	-72	21	0.22	L Middle Temporal Gyrus; Area Pgp (IPL) 56%
		-57	-45	27	0.17	L SupraMarginal Gyrus; Area Pfcm (IPL) 40%
		-39	-84	9	0.16	L Middle Occipital Gyrus; Area hOc4lp 59%
		-36	-72	36	0.15	L Middle Occipital Gyrus
		-36	-87	15	0.14	L Middle Occipital Gyrus; Area hOc4lp 80%
R Cerebellum (VIIb,IX,X)	75	24	-42	-45	0.12	R Cerebellum (VIII)
		15	-42	-45	0.12	R Cerebellum (IX)
		12	-48	-48	0.11	R Cerebellum (IX); Lobule IX (Hem) 59%
R Middle Frontal Gyrus	65	45	3	54	0.20	R Middle Frontal Gyrus
L Middle Frontal Gyrus L BA44	65	-48	12	45	0.19	L Middle Frontal Gyrus
		-36	9	27	0.15	L IFG (p. Opercularis)
		-48	0	36	0.11	L Precentral Gyrus
		-45	0	27	0.10	L Precentral Gyrus
L Insula	44	-36	24	3	0.16	L Insula Lobe
		-36	15	-9	0.12	L Insula Lobe
		-33	21	-6	0.11	L Insula Lobe
L Amygdala	39	-15	3	-12	0.11	N/A
		-18	-3	-12	0.11	L Amygdala; BF (Ch 4) 53%
		-18	-6	-21	0.11	L Hippocampus; Amygdala (LB) 49%
		-18	15	-6	0.11	L Putamen
		-21	0	-15	0.10	L Amygdala
L BA45	37	-51	33	12	0.18	L IFG (p. Triangularis); Area 45 66%
R Superior Frontal Gyrus	34	21	15	60	0.19	R Superior Frontal Gyrus
		24	18	51	0.16	R Superior Frontal Gyrus
Medial Recentral Gyrus	19	-3	-21	66	0.15	L Paracentral Lobule
		3	-21	69	0.12	R Posterior-Medial Frontal
third memorize>first memorize ∩ third memorize>0						
L Medial Prefrontal Cortex	26	-6	60	24	0.16	L Superior Medial Gyrus
		-6	54	33	0.15	L Superior Medial Gyrus
		-3	57	12	0.13	L Superior Medial Gyrus; Area Fp2 59%
		-6	54	15	0.13	L Superior Medial Gyrus
Medial Orbitofrontal Cortex	21	-3	36	-15	0.19	L Rectral Gyrus; Area s32 52%
R Cerebellum (Crus1+2)	17	30	-87	-36	0.13	R Cerebellum (Crus 2); Lobule VIIa crusl (Hem) 70%
		18	-90	-36	0.12	R Cerebellum (Crus 2); Lobule VIIa crusII (Hem) 94%

Table 3.3: Synchrony changes between the first and third memorize session. 'stat.' denotes Fisher z-values. Corresponding table to figure 3.3b. q<0.001 FDR corrected.

location	cluster size	peak statistics x	y	z	stat.	peak location *area with max probability*
first memorize>count folds ∩ first memorize>0						
R+L Cerebellum	5403	-3	-81	2	0.27	L Lingual Gyrus; Area hOc1 [V1] 41%
(VI,V,Crus1+2,VIII,IX)		57	-30	30	0.25	R Supra Marginal Gyrus; Area Pfcm (IPL) 29%
Occipital Cortex, IPL,SII		9	-78	0	0.25	R Lingual Gyrus; Area hOc1 [V1] 57%
		9	-75	-6	0.23	R Lingual Gyrus; Area hOc2 [V2] 43%
		60	-57	15	0.22	R Middle Temporal Gyrus; Area Pga (IPL) 63%
		-45	-51	-12	0.22	L Inferior Temporal Gyrus; Area FG4 83%
		-51	-60	-15	0.22	L Inferior Temporal Gyrus; Area FG2 35%
		-54	-69	12	0.21	L Middle Temporal Gyrus
		-33	-66	-21	0.21	L Cerebellum (VI); Area Lobule VI (Hem) 38%
		-60	-60	9	0.21	L Middle Temporal Gyrus
		66	-24	39	0.21	R Supra Marginal Gyrus; Area PF (IPL) 29%
R+L SPL (Area 5)	558	18	-54	72	0.18	R Superior Parietal Lobule; Area 5L (SPL) 52%
Area 2		-3	-42	69	0.16	L Precuneus; Area 4a 33%
		-3	-48	66	0.16	L Precuneus; Area 5M (SPL) 52%
		3	-48	66	0.16	R Precuneus; Area 5M (SPL) 48%
		21	-51	69	0.16	R Superior Parietal Lobule; Area 5L (SPL) 49%
		9	-48	39	0.16	R Precuneus
		3	-48	48	0.15	R Precuneus
		24	-48	72	0.15	R Superior Parietal Lobule; Area 7PC (SPL) 55%
		30	-45	69	0.15	R Postcentral Gyrus; Area 2 30%
Anterior Cingulate	201	-3	24	39	0.19	L Superior Medial Gyrus
Anterior Paracingulate		3	27	36	0.17	R MCC
		0	6	39	0.15	L MCC
Middle Temporal Gyrus	174	51	3	-21	0.15	R Middle Temporal Gyrus
Superior Temporal Sulcus		45	18	-27	0.14	R Temporal Lobule
		51	-15	-12	0.14	R Middle Temporal Gyrus
R+L Thalamus	97	-12	-33	-3	0.12	L Lingual Gyrus
		9	-33	3	0.12	N/A
		12	-30	-3	0.12	N/A
		-21	-27	-9	0.12	L Hippocampus
		27	-21	-9	0.11	Thal: Temporal 44%
R Area 45	96	51	39	9	0.21	R IFG (p. Triangularis)
DLPFC		54	27	0	0.14	R IFG (p. Triangularis); Area45 33%
Medial SPL	91	6	-72	39	0.14	R Precuneus
Area7		-9	-66	33	0.14	L Precuneus
		18	-84	36	0.14	R Superior Occipital Gyrus; Area hOc4d [V3A] 43%
L BA44	54	-54	6	21	0.12	L Precentral Gyrus; Area 44 58%
		-51	6	9	0.12	L IFG (p. Opercularis)
		-36	6	27	0.12	L IFG (p. Opercularis)
L BA45, DLPFC	49	-48	33	12	0.18	L IFG (p. Triangularis); Area 45 46%
R Precentral Gyrus	47	48	3	51	0.18	R Precentral Gyrus
L Precentral Gyrus	45	-42	3	54	0.15	L Precentral Gyrus
		-48	-6	54	0.13	L Precentral Gyrus
Thalamus	43	3	-15	12	0.14	R Thalamus; Thal: Temporal 70%
R BA44	41	48	15	27	0.14	R IFG (p. Opercularis)
		45	9	21	0.14	R IFG (p. Opercularis)
L Amygdala	31	-21	6	-12	0.11	N/A
		-21	-6	-15	0.10	L Amygdala
		-30	-3	-15	0.10	L Amygdala
		-18	-6	-21	0.09	L Hippocampus; Amygdala (LB) 49%
L Caudate	27	-6	6	3	0.13	L Caudate Nucleus
R V1	19	6	-57	6	0.11	R Lingual Gyrus; Area hOc1 [V1] 46%
count folds>first memorize ∩ count folds>0						
R V2	21	18	-99	18	0.18	R Superior Occipital Gyrus; Area hOc2[V2] 63%

Table 3.4: Synchrony changes between the first memorize session and the count folds session. Corresponding table to figure 3.3b. 'stat' denotes Fisher z-values. q<0.001 FDR corrected.

3.3.4 Relation of the common component to the optic flow of the videos

The visual motion of the videos is completely generated by the instructive hand movements. We wanted to investigate if we can relate this visual motion of the video to the common observer activity. Therefore, we correlated the optic flow with the first principal component. Figure 3.4 shows the significant correlations of the optic flow of the butterfly video with the first principal component of the first memorize run, and the correlation of the control video with the first principal component of the 'count folds' run. Overall we found high correlations, i.e.a high level of similarity, between the optic flow of the video and the time course/score of the first principal component of the observers during observation in the AOEN. We found high correlations not only in motion processing areas like V5 and parietal ares but also in the primary somatosensory cortex, right dorsal and ventral premotor cortex and cerebellar regions including lobules VI and VIII a and b. Correlations ranged between 0.4 and 0.8 for cortical areas. Moreover, we found significant high anti-correlations in primary motor areas (for the memorize run), IPL to angular gyrus (both) and cuneal areas (for the count folds run), ranging between -0.4 and -0.7. Exact locations and correlation values are depicted in tables 3.5 and 3.6. Figure 3.5 shows two exemplary time courses, one for a somatosensory voxel, and one for a motor voxel. Even though significance was reached for partly different areas when visually inspecting the two conditions of the experiment, the unthresholded data shows that both conditions have a similar behavior subthreshold (see supplementary figure 3.9).

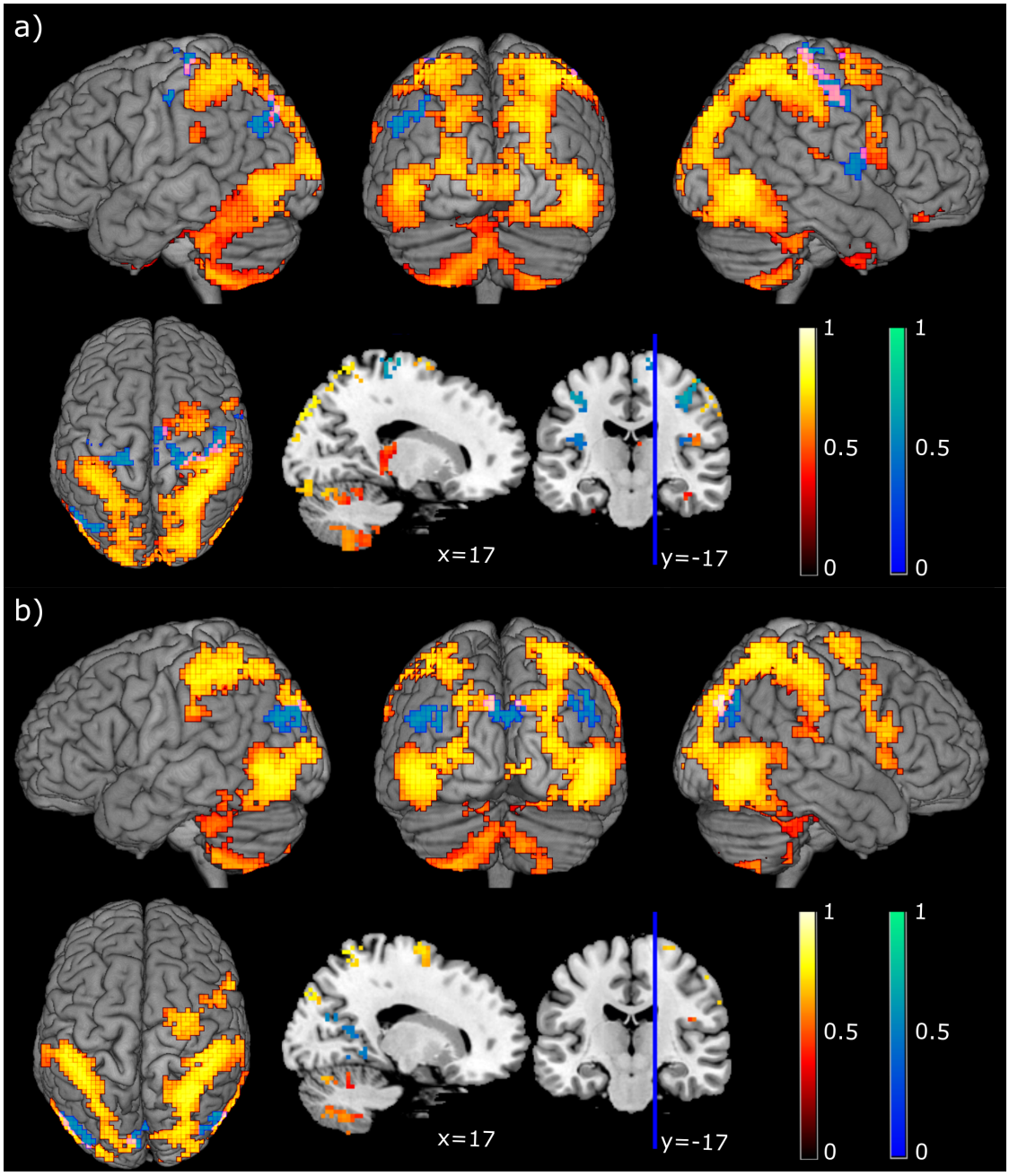


Figure 3.4: Significant correlations of the time course of the first component and the optic flow of the corresponding video. The optic flow correlates highly with the first component in action observation areas. Thresholded at $q<0.001$ FDR corrected, $k\geq19$. Note that there is no real overlap between positive and negatively correlated areas; the seen overlap is a result of the rendering process. a) correlation with first watch to memorize condition (butterfly video). b) correlation with control condition (unproductive video).

location	cluster size	peak statistics x	y	z	corr.	peak location *area with max probability*
positive correlations of first memorization run and optic flow						
R+L Cerebellum (VI, VIII), Occipital Cortex, SPL, IPL, Postcentral Gyrus	2266	-18	-100	14	0.79	L Middle Occipital Gyrus; Area hOc3d [V3d] 37%
		-48	-85	2	0.77	L Middle Occipital Gyrus; Area hOc4la 83%
		-42	-49	65	0.77	L Inferior Parietal Lobule; Area 2 35%
		-45	-40	59	0.75	L Postcentral Gyrus; Area 2 70%
		-48	-43	62	0.75	L Postcentral Gyrus; Area 2 52%
		-27	-55	74	0.75	L Superior Parietal Lobule; Area 7A (SPL) 31%
		-39	-79	2	0.75	L Inferior Occipital Gyrus; Area hOc5 [V5/MT] 54%
		-27	-52	-52	0.75	L Cerebellum (VIII); Lobule Villa (Hem) 61%
R Cerebellum (VI), Occipital Cortex, SPL, IPL, Postcentral Gyrus	2059	51	-70	2	0.84	R Inferior Temporal Gyrus; Area hOc4la 38%
		33	-61	68	0.83	R Superior Parietal Lobule; Area 7PC (SPL) 37%
		42	-43	65	0.82	R Postcentral Gyrus; Area 1 40%
		42	-37	65	0.82	R Postcentral Gyrus; Area 1 45%
		27	-64	71	0.81	R Superior Parietal Lobule; Area 7A (SPL) 74%
		24	-91	44	0.79	R Superior Occipital Gyrus; Area hOc4d [V3A] 40%
		30	-52	74	0.79	R Postcentral Gyrus; Area 2 26%
		18	-91	44	0.79	R Cuneus; Area hOc4d [V3A] 27%
		39	-37	47	0.78	R SupraMarginal Gyrus; Area 2 48%
R Precentral Gyrus, Superior Frontal Gyrus	98	36	-7	68	0.73	R Superior Frontal Gyrus
		24	-10	77	0.70	R Superior Frontal Gyrus
		30	-10	56	0.67	R Precentral Gyrus
		24	5	71	0.63	R Superior Frontal Gyrus
R Fusiform Gyrus, Entorhinal Cortex	69	33	-7	-40	0.49	R Fusiform Gyrus
		30	2	-40	0.47	R Inferior Temporal Gyrus; Area hOc4la 38%
		36	-10	-31	0.45	R Fusiform Gyrus
R IFG	65	54	8	41	0.70	R Precentral Gyrus
		54	8	23	0.61	R IFG (p. Opercularis); Area 44 65%
		48	5	20	0.60	R IFG (p. Opercularis); Area 44 35%
R Thalamus (temporal, parietal)	62	15	-34	2	0.54	N/A; Thal Visual 25%
		18	-34	17	0.54	R Thalamus; Thal Temporal 69%
		6	-22	20	0.45	R Thalamus; Thal Temporal 79%
R Operculum	25	39	-7	20	0.51	R Rolandic Operculum; Area OP3 [VS] 34%
		42	-13	23	0.45	R Rolandic Operculum; Area OP3 [VS] 59%
L IPL, Supramarginal Gyrus	22	-63	-34	32	0.63	L SupraMarginal Gyrus; Area PF (IPL) 33%
		-57	-34	35	0.58	L SupraMarginal Gyrus; Area Pft (IPL) 34%
R SII, Suparamarginal Gyrus	21	48	-25	20	0.65	R Rolandic Operculum; Area OP1 [SII] 64%
		57	-28	23	0.54	R SupraMarginal Gyrus; Area OP1 [SII] 44%
R IFG	21	36	29	-16	0.63	R IFG (p. Orbitalis)
negative correlations of first memorization run and optic flow						
R Precentral Gyrus, Area 4a	75	51	-16	56	-0.66	R Precentral Gyrus
		39	-22	53	-0.66	R Precentral Gyrus; Area 4a 23%
		48	-22	65	-0.65	R Precentral Gyrus
R Operculum/Insula	64	54	-7	14	-0.60	R Rolandic Operculum
		39	-16	23	-0.48	R Insula Lobe; Area OP3 [VS] 60%
R Precentral Gyrus, Area 4a	53	21	-31	80	-0.72	R Precentral Gyrus; Area 4a 39%
		27	-31	68	-0.70	R Postcentral Gyrus
		18	-25	80	-0.68	R Precentral Gyrus
L Postcentral Gyrus, Area 4a	42	-24	-31	71	-0.68	L Postcentral Gyrus; Area 4a 35%
		-36	-31	74	-0.67	L Postcentral Gyrus; Area 4a 52%
		-18	-31	77	-0.66	L Postcentral Gyrus; Area 4a 50%
		-18	-25	80	-0.59	L Paracentral Lobule
L Operculum/Insula	23	-45	-25	20	-0.52	L Rolandic Operculum; Area OP1 [SII] 54%
		-36	-22	23	-0.49	L Insula Lobe; Area OP3 [VS] 53%
R Paracentral Lobule, Area 4a	21	3	-16	68	-0.61	R Posterior-Medial Frontal
		9	-13	77	-0.60	R Posterior-Medial Frontal
		6	-10	74	-0.59	R Posterior-Medial Frontal
		3	-31	68	-0.58	R Paracentral Lobule; Area 4a 79%
L IPL/Angular Gyrus	20	-42	-82	44	-0.66	L Middle Occipital Gyrus; Area Pgp (IPL) 59%
		-51	-73	41	-0.66	L Angular Gyrus; Area Pgp (IPL) 48%
L Postcentral Gyrus, Area 4p	19	-39	-22	53	-0.63	L Postcentral Gyrus; Area 4p 40%

Table 3.5: Correlation between the first principal component/score of the first memorize run and the optic flow of the video. Table corresponds to figure 3.4a. q<0.001 FDR corrected, k≥ 19.

location	cluster size	peak statistics x	y	z	corr.	peak location *area with max probability*
positive correlations of count folds run and optic flow						
R Occipital Cortex, SPL, IPL, SII, Postcentral Gyrus	1519	51	-73	2	0.85	R Inferior Temporal Gyrus; Area hOc4la 68%
		42	-40	59	0.83	R Postcentral Gyrus; Area 2 56%
		42	-37	65	0.83	R Postcentral Gyrus; Area 1 45%
		27	-64	71	0.82	R Superior Parietal Lobule; Area 7A (SPL) 74%
		33	-54	63	0.82	R Superior Parietal Lobule; Area 7PC (SPL) 48%
		27	-82	41	0.79	R Superior Occipital Gyrus
		33	-52	56	0.78	R Inferior Parietal Lobule; Area 7PC (SPL) 21%
		30	-82	50	0.78	R Superior Occipital Gyrus
L Occipital Cortex, Cerebellum VI	672	-54	-76	-1	0.83	L Inferior Occipital Gyrus; Area hOc4la 82%
		-45	-85	2	0.81	L Inferior Occipital Gyrus; Area hOc4la 92%
		-39	-79	2	0.81	L Inferior Occipital Gyrus; Area hOc5 [V5/MT] 54%
		-42	-67	5	0.80	L Middle Occipital Gyrus
		-24	-94	14	0.80	L Middle Occipital Gyrus; Area hOc4lp 43%
		-36	-85	8	0.77	L Middle Occipital Gyrus; Area hOc4la 53%
		-21	-88	20	0.75	L Middle Occipital Gyrus
		-39	-79	8	0.75	L Middle Occipital Gyrus; Area hOc5 [V5/MT] 38%
		-45	-46	-10	0.70	L Inferior Temporal Gyrus; Area FG4 63%
R+L Cerebellum (VIII, VIII, IX, X)	470	-30	-46	-46	0.72	L Cerebellum (VII); Lobule VIIIa (Hem) 71%
		-24	-58	-46	0.71	L Cerebellum (VII); Lobule VIIIb (Hem) 48%
		-9	-76	-37	0.70	L Cerebellum (VII); Lobule VIIIa (Hem) 52%
		12	-82	-40	0.67	R Cerebellum (VII); Lobule VIIb (Hem) 53%
		24	-61	-46	0.65	R Cerebellum (VIII); Lobule VIIb (Hem) 41%
		27	-46	-46	0.63	R Cerebellum (VIII); Lobule VIIIb (Hem) 71%
		-39	-67	-49	0.62	L Cerebellum (VII); Lobule VIIb (Hem) 61%
		3	-75	-28	0.57	Cerebellar Vermis (7); Lobule VIIb (Verm) 56%
L Postcentral Gyrus, IPS, SPL	401	-48	-34	59	0.82	L Postcentral Gyrus; Area 2 38%
		-39	-46	56	0.80	L Inferior Parietal Lobule; Area 2 25%
		-42	-49	59	0.80	L Inferior Parietal Lobule; Area 7PC (SPL) 21%
		-36	-58	62	0.79	L Inferior Parietal Lobule; Area 7A (SPL) 52%
		-30	-67	59	0.78	L Superior Parietal Lobule
		-42	-37	68	0.77	L Postcentral Gyrus; Area 1 57%
R Cerebellum (V, VI)	161	21	-76	-16	0.68	R Cerebellum (VI); Lobule VI (Hem) 92%
		27	-55	-19	0.67	R Cerebellum (VI); Lobule VI (Hem) 90%
		21	-61	-13	0.60	R Cerebellum (VI); Lobule VI (Hem) 95%
		33	-40	-22	0.54	R Cerebellum (IV-V); Area FG4 43%
		30	-34	-25	0.49	R Cerebellum (IV-V)
R Precentral Gyrus, Superior Frontal Gyrus	135	27	-10	74	0.76	R Superior Frontal Gyrus
		27	-16	77	0.76	R Precentral Gyrus
		33	-7	68	0.76	R Superior Frontal Gyrus
		30	-10	65	0.75	R Superior Frontal Gyrus
		18	-7	74	0.72	R Superior Frontal Gyrus
R IFG	94	60	14	23	0.76	R IFG (p. Opercularis); Area 45 64%
		57	11	29	0.74	R IFG (p. Opercularis); Area 44 39%
		48	2	44	0.72	R Precentral Gyrus
L Occipital Cortex	61	-21	-91	44	0.77	L Superior Occipital Gyrus; Area hOc4d [V3A] 48%
		-9	-97	33	0.76	L Cuneus; Area hOc3d [V3d] 57%
		-6	-94	41	0.75	L Cuneus; Area hOc3d [V3d]
		-18	-94	38	0.74	L Superior Occipital Gyrus; Area hOc4d [V3A] 66%
L Cerebellum (VI)	36	-18	-73	-19	0.66	L Cerebellum (VI); Lobule VI (Hem) 99%
L IPL, Supramarginal Gyrus	28	-60	-31	38	0.70	L SupraMarginal Gyrus; Area Pft (IPL) 60%
R Occipital Cortex	22	15	-97	5	0.79	R Calcarine Gyrus; Area hOc1 [V1] 68%
		9	-94	2	0.78	R Lingual Gyrus; Area hOc1 [V1] 78%
negative correlations of count folds run and optic flow						
R+L Cuneus	232	-3	-94	38	-0.71	L Cuneus; Area hOc3d [V3d] 71%
		-6	-88	44	-0.68	L Cuneus; Area 7P (SPL) 31%
		6	-76	32	-0.64	R Cuneus
		-12	-58	11	0.57	L Calcarine Gyrus
		21	-73	26	-0.55	R Cuneus
L IPL	33	-48	-82	32	-0.68	N/A; Area Pgp (IPL) 68%
		-36	-91	32	-0.64	L Middle Occipital Gyrus; Area Pgp (IPL) 42%
R IPL, Angular Gyrus	19	48	-69	39	-0.68	R Angular Gyrus; Area Pgp (IPL) 78%
		42	-82	41	-0.60	R Middle Occipital Gyrus; Area Pgp (IPL) 74%
		45	-82	35	-0.57	R Middle Occipital Gyrus; Area Pgp (IPL) 80%

Table 3.6: Correlation between the first principal component/score of the count folds run and the optic flow of the video. Table corresponds to figure 3.4b. q<0.001 FDR corrected, k≥ 19.

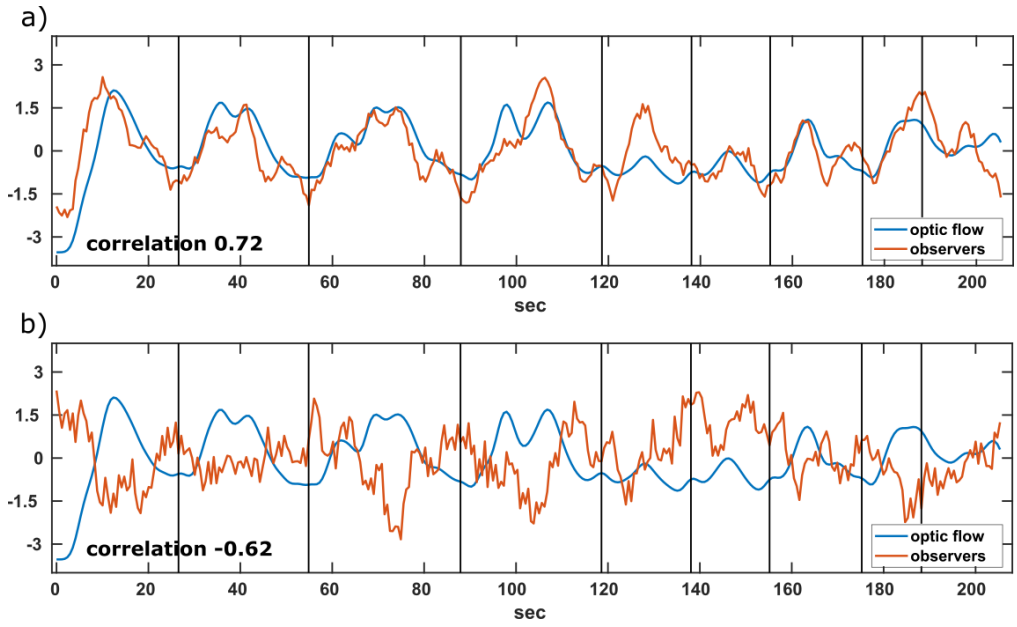


Figure 3.5: The optic flow time course (HRF-convolved, blue) for the first memorize video plotted together with time course of the first component/score for exemplary voxels (red). The optic flow time course follows the time course of the first principal component. a) somatosensory voxel -33,-42,63. b) motor voxel -20,-27,62. (For both voxels, the first component shows a significant explained variance.)

3.4 Discussion

The aim of this paper was to investigate the initial learning stage during naturalistic observational skill learning. Special emphasis was placed on developing a paradigm that enabled naturalistic observation of instructive videos and action reproduction sequentially completely inside the MRI machine. By using data-driven analysis methods like the PCA or the video feature analysis of the visual optic flow we were able to determine new aspects of the AOEN and observational skill learning while keeping a naturalistic setup. We demonstrate a consistent synchrony of the AOEN during action observation across the first attempts of reproduction as well as pure action identification. The synchrony between the observers decreases between the initial learning and the end of the learning session; it is also higher for the initial learning run than for the count folds run. Additionally to these learning-specific results, we show that the visual optic flow, putatively a proxy for hand motion, is highly represented in the AOEN well beyond visual areas. Moreover, our data suggests that additional areas should be included into the AOEN: cerebellar areas VI and VIIIa/b as well as the primary motor cortex.

3.4.1 Involvement of the AOEN and beyond

Consistent between-subjects similarity As Thomas et al. (2018) have shown, when observing sequences of actions, classical AOEN areas are synchronized across viewers. We can reproduce these findings: we find a consistent common principal component for all classical AOEN areas, namely the IPL, SPL, somatosensory cortex, dPMC and vPMC and SMA. This means the observers' activity is synchronized across a range of different cognitive tasks taking place during action observation: We find consistent synchrony for learning to reproduce a sequence of actions, for monitoring own actions (when the subjects had already reproduced that step) and for identifying actions as in the counting folds run. Whether the displayed sequence was goal-directed (in the sense of producing a final origami figure) was not relevant for the overall similarity in these AOEN areas since the instructive video for the counting folds run did not produce a paper figure. These findings support the universal involvement of the AOEN in diverse action observation tasks.

Additionally we would like to stress the involvement of not-so established areas, in particular cerebellar areas. The consistently similar areas across all observation tasks are in a similar location to the motor execution areas of complex hand or finger movements (Schlerf et al., 2010; Mottolese et al., 2013), but the synchrony also stretches beyond this. Importantly, our results fit well with the findings of the recent preprint of Abdelgabar et al. (2018) of cerebellar involvement during action observation. Consequently, with this work we would also like to establish the cerebellar areas VI and VIIIa and VIIIb as

parts of the AOEN.
Another interesting aspect is the involvement of apparent primary motor areas. While the variance explained by the first component is low, it is still significant. We will refer more to this finding in section 3.4.3.

Within-subject similarities Some studies on temporal similarities during video observation have also investigated a synchrony across sessions within subjects, termed intrasubject correlation by Golland et al. (2007). They have shown a high similarity for repeated viewings of videos (parts of feature movies, see Golland et al., 2007; Nguyen et al., 2017, or a head shot of a person explaining tasks Andric et al., 2016). In our case, for the repeated viewings of the memorization video, we have found high intrasubject correlation in visual areas, AOEN areas and cerebellar areas. This indicates that the subjects employ the AOEN very similarly during the repeated viewings, even though the cognitive tasks vary: from pure memorization in the beginning to mainly monitoring one's own steps in the third run. Moreover, arguably attentiveness may be lower due to the repeated viewings. Still, our results show a high similarity of activation within each subject in the AOEN. This stresses the universal nature of the AOEN in the processing of action stimuli, irrespective of the task. Along with the findings that the AOEN is robustly synchronized between subjects for each run, our results of processing similarity within each subject further demonstrate the robustness of the AOEN as an 'always-there' system during action observation.
On the other hand, we also find significant intrasubject correlations for the first learning run and the counting folds run. These two runs have been counterbalanced during the experiment so that order effects can be ruled out. Here it is important to note that the videos for both tasks are different in content, but still share some basic features, such as the common setup or similarities in movements. In this case we still find within-subject similarities in the motion sensitive areas like V5/MT, the SPL, and a small part of the premotor system, although showing low correlations. Consequently, the processing of similar (but in detail different) stimuli during action observation shows a commonness of activation in motion sensitive areas and parts of the AOEN within each subject, albeit low if the correlation values are regarded.
Taken together, we can conclude that the AOEN is a robust system generally involved in the processing of observed actions. However, the detailed processing within each subject is largely specific to the content of the visual stimulus. Nevertheless, low-level commonalities are still to a certain low extent processed similarly.

Changes in synchrony As we can compare the similarity between sessions we can also compare the differences in synchrony between them. For these synchrony changes, the highest correlation changes happened in the lateral occipito-temporal cortex (LOTC)

and TPJ. The LOTC is involved in a diverse number of cognitive tasks, ranging from semantic tasks like verb (in contrast to noun) processing to action observation to the perception and inference of tool use (Lingnau and Downing, 2015). As a part of this, the posterior part of the superior temporal sulcus (pSTS) has been reported to show higher activation for imitation than for passive observation (Molenberghs et al., 2010) and higher activity for observation with subsequent imitation than observation with a subsequent self-selected motor task (Mainieri et al., 2013). These findings are in line with lower synchrony for the counting condition as no imitation is required here. A possible hypothesis for both findings would be that the pSTS is especially involved in the imitation of new content, correlating potentially with attentional load. As in the third learning run, most people have almost completely memorized the origami and therefore mostly monitor their steps. Also, some subjects have reported in the debriefing not having paid much attention for the third run.

The reported areas stretch to the TPJ and supramarginal gyrus. In action observation studies, classically this area has been associated with mentalizing (see e.g. Van Overwalle and Baetens, 2009). However, we would like to argue that classical mentalizing in the sense of actively inferring intentions from the instructor during the high-load learning condition is unlikely. Schippers et al. (2009) argue that the TPJ typically coactivates with the medial prefrontal cortex in mentalizing tasks, which is also not the case here. As an alternative explanation, they mention attentional reorienting. Carter and Huettel (2013) review the TPJ as a nexus area showing common or bordering activity in attentional, memory, language and theory of mind tasks. They argue for especially the supramarginal gyrus as a convergence area for the dorsal and ventral attention stream and the memory stream, but also for the complete TPJ as a nexus area incorporating all information to infer social situations. For our data, the hypothesis of a connector between attention circuits and memory formation would be fitting, as in comparison to the control condition, no memory (except for low-load working memory) is required and arguably less attentional demand. The latter also would be true for the third learning run; also less active encoding would have to be done here.

Consistently with the findings on imitation by Buccino et al. (2004) and Higuchi et al. (2012), we have also found higher synchrony in the DLPFC for imitation learning than counting, albeit our clusters are partly inferior to theirs. They have reported the DLPFC specifically for immediate imitation and not necessarily for observational tasks (Higuchi et al., 2012).

We have also found that large parts of the cerebellum are more strongly correlated in both comparisons, also beyond classical motor areas. This poses an interesting fact for the involvement of the cerebellum in action observation research, as, depending on the task, different non-motor areas show a change in synchrony here. When comparing the

results to the network connectivity structure presented in Marek et al. (2018) we find overlap with the cingulo-opercular and fronto-parietal networks, which are known as attentional regulation and control networks. Consequently, one possible explanation of the findings of similar higher similarity for the first memorize run could be higher attentional and control demands, which would arise from a potentially higher cognitive load. The maps also show a good overlap with the 'memorize a spatial map' task (for 'medium' and 'hard' difficulty) of the recent preprint of King et al. (2019) and are also similar to the results of Stoodley et al. (2012) and Küper et al. (2016) for visuo-spatial working memory. Memorizing spatial features has certainly been a cognitive task performed in the first memorize run. It is not needed for counting folds, where actions need to be identified, but only one number has to be kept in working memory. For the third memorize run, most parts are already encoded, which is also consistent with the finding.

3.4.2 Optic flow as a putative proxy for hand movement

Raudies (2013) defines optic flow as 'the change of structured light in the image [...] due to relative motion between the eyeball or camera and the scene'. In the case of our videos, changes of brightness between image frames are calculated as a vector field known as the optic flow. For movies, the total change of luminance can be separated into global flow, local flow and residual pure illumination changes (Bartels et al., 2008). Global flow or global motion is elicited by camera movements like pans or zooms and could be compared to self-motion relative to the scene (Bartels et al., 2008; Dayan et al., 2018). On the other hand, local motion depicts the movements of objects or persons in the scene. Cuts would also elicidate luminance changes but are cut out in the original analysis of Bartels et al. (2008). In our case, the camera perspective is fixed, non-moving, and no cuts are employed. (In the routine Bartels et al. (2008) describe, global motion computation can be regarded as mathematically equivalent to a regression fitting optic flow to ideal full translation, expansion and rotation fields.) We argue therefore that, by calculating the optic flow with a classical algorithm, we compute the local flow.
We do find classical motion sensitive areas previously linked to optic flow in general and also local motion correlating strongly also in our cases, namely area MT/V5, classically linked to motion processing or area V3A. We have also found correlations in areas linked to global motion processing like the posterior parietal cortex (Bartels et al., 2008) or the right IFG (Dayan et al., 2018). Nevertheless we also find high correlations in classical AOEN areas like broader areas of the parietal cortex (IPL, SPL) or the somatosensory cortex (see also fig. 3.5), and, additionally, cerebellar AOEN areas. Given that the optic flow depicts motion in videos, we argue that these results show the processing of motions. However, to our knowledge these areas have not been reported yet. The reason could be twofold: as classical feature movie pieces have been used, local motion is normally not that much focussed on by the participants and appears less isolated even

in the calculation of local motion. In our case, local motion is the only input and must be actively processed in both tasks (must be identified for memorizing it or for counting it). Consequently, the process might be masked by less attention/active processing and less clear optic flow. Additionally, perspective might matter: while only global motion has been associated with self-motion, we observe correlated areas that have only been associated with global flow. However, both videos were shot in ego-perspective, which is normally not used in feature movies. Consequently, this input might be computed in a way more similar to self-motion than the third-person perspective.

On the other hand, one can argue that the optic flow is a proxy for hand movement: The only objects moving in the video are the sheet of paper and the pair of hands. Additionally, the sheet of paper movement is completely generated by and therefore dependent on the hand movements. Supporting this argument, we have correlated the instructor's activity with the optic flow and have found correlations ranging from 0.4 to 0.5 in parts of the AOEN, particularly parts of the somatosensory and premotor cortex. See supplementary figure 3.9 a) and b) for correlation maps. (Please note that the displayed data are uncorrected and therefore presented merely for illustrative purposes.) As the instructor is blindfolded, optical properties are irrelevant to her processing. However, of course it is to be assumed that the instructor's activity in the AOEN is directly related to the motoric output she is producing. Consequently, it is highly likely that the correlation between the instructor's activity and the optic flow arises due to a high correlation between the optic flow and the motor actions. We therefore interpret the optic flow of the videos as a proxy for the shown hand movements and the hand-paper manipulation. Along with this, correlations between the observers' first principal component and the optic flow are high in the somatosensory cortex (with correlations of up to 0.8), suggesting that the optic flow has a big explanatory power for the common activation structure of all participants in this area, even without the specific somatosensory input.

Interestingly, we find strong anti-correlations alongside the central sulcus, particularly in BA3 and 4, which poses parts of the primary somatosensory cortex and primary motor cortex. While this is not a common finding in the AOEN literature, primary motor areas have been reported to be down-regulated during action observation (Gazzola and Keysers, 2009). We will reflect more on the M1 involvement in 3.4.3.

Taken together, our findings offer the following contributions: First of all, we can associate the processing of the optic flow with the neural activity of areas previously not reported to be associated with optic flow; particularly premotor and somatosensory areas. However, it has to be kept in mind that we present correlative data. Consequently, it is possible that this is driven by a third correlated variable, in our case hand motion. The great effect size and the fact that the observer input was merely visual would suggest a direct processing of optical stimuli; however, the effect could also be generated by

the mapping of the incoming stimuli onto the own motor system. This notion would be supported by the associations between the instructor activity and the optic flow we show in fig. 3.9 even though the instructor is blindfolded.
If optic flow can be regarded as a proxy for hand motion (as every motion in the video is hand-generated), in the long run this offers other possibilities for automatic image processing of actions in fMRI: One of the problems of intersubject correlation between instructor and observers is that the stimuli have match exactly time-wise. However, human performance itself is variable. Previously several subjects have been matched upon each other based on fine-grained annotation of the videos of the action production and dilating/stretching in between (by performing a so-called time-warp, Silbert et al., 2014). However, annotation is always tedious and subjective. If a parameter like the optic flow offers a good correlation in certain voxels, these (and therefore the whole brain, if we assume time-wise similar processing) could be matched automatically between subjects, effectively warping subjects onto each other.

3.4.3 M1 involvement

In the central sulcus we have found significantly correlated activity between subjects (significant first principal component) and high anti-correlations of the first principal component to the video's optic flow. We have to take into account here the resolution of 3x3x3 mm and our usage of the DARTEL algorithm, the first aim of which is to match the subjects to each other by creating a common template. The resulting areas include parts of BA 4a and p as well as BA2, which would be a common cluster in the primary motor cortex and primary somatosensory cortex and which also fits the visual inspection of the data.
To our knowledge, M1 deactivation in fMRI during action observation has only been shown by Gazzola and Keysers (2009); intersubject correlation in the respective area is shown by Thomas et al. (2018) but not discussed in further detail. Typically, activation-based studies on the AOEN do not report M1 involvement.
Nevertheless, the mirror neuron literature investigating non-human primates reports neurons with 'mirror-properties' in M1 (for a review of mirror neurons in non-human primates, see Kilner and Lemon, 2013). In particular, Dushanova and Donoghue (2010) and Vigneswaran et al. (2013) report neurons with 'mirror behavior' in M1; Kraskov et al. (2014) report M1 pyramidal tract neurons with mirror suppression behavior. These findings show not only neurons with the classical mirror behavior (active during action execution and the observation thereof) but also with a suppression behavior, meaning an activation during action execution but a suppression during action observation. Both types are found in M1. The authors of these studies suggest that the suppression mirror neurons play an important role in regulation during action observation; possibly this

suppressive character is an explanation for why the motor system activates during action observation, but does not produce the same action at the same time (as even pyramidal tract neurons show the classical mirror neuron behavior (Kraskov et al., 2009, 2014)).
This offers a possible explanation of our data: We find a common component in M1 during observation due to similar behavior of putative mirror neurons. This does not conclude an up- or down-regulation of activity as we look for similarity between subjects, without a hypothesis about magnitude changes. The explained variance of the component is nevertheless relatively low - which shows the diverse ongoing processes in this area (see e.g. activation as well as suppression of putative neurons with mirror behavior). If the optic flow is a proxy for hand movement, our findings suggest a down-regulation of activity whenever motion or performed action is high. This is consistent with the finding of suppression mirror neurons in M1 during action observation in monkeys, under the assumption that these neurons exist also in humans. Nevertheless, the findings of down-regulation/anti-correlation can also be explained by a putative suppression mechanism in M1 during action observation which does not have to be generated by mirror neurons. Further research is needed to investigate this behavior.

3.4.4 Conclusion and outlook

With our study we broaden the knowledge of the AOEN by reporting a commonness in activation during naturalistic action identification and action learning. The classical AOEN is a system that is robustly synchronized across and within subjects during action observation, across a diverse range of cognitive tasks. Following these lines, we can also report a consistent and robust involvement of the cerebellum and especially cerebellar hand motion areas; not only for the learning process but also for the process of identifying actions, also without a goal-directed context. This stresses the importance of the previously neglected function of the cerebellum in cognition, and not only in learning contexts. Our results demonstrate that the cerebellum should be regarded as part of the AOEN. Additionally, we show an involvement of the primary motor cortex during action observation; also in a non-learning task of action identification. We strongly encourage further research in humans in this context.
We can demonstrate that the between-observers synchrony decreases during the initial first attempts of learning, arguably because of reduced attentional and especially reduced cognitive control demands, and less active memory encoding. Similarly, we demonstrate that the between-observers synchrony is higher for the memorization than the action identification task, which can also be explained by reduced cognitive control demands and less active memory encoding. Moreover, we also present the PCA algorithm as an extension to the classical intersubject correlation approach.
Last but not least, we show that the optic flow of an action video is strongly associated

with diverse action processing areas of the brain which have previously not been associated with optic flow processing. We postulate that the optic flow of a video depicting actions is a good proxy for hand movements and may therefore expand the possibilities of automated analysis of video and MRI data. This way, it might enable different naturalistic motor and movement experiments.

Taken together, the naturalistic paradigm and the data-driven analysis methods open up new possibilities of researching real-life processes. Understanding observational video learning will be of great importance for the future trend of e-learning. Since naturalistic paradigms like the one presented in this study are easily understood, they can be adapted to different subject and patient groups to investigate action processing of naturalistic stimuli in health and disease.

3.5 Supplementary material

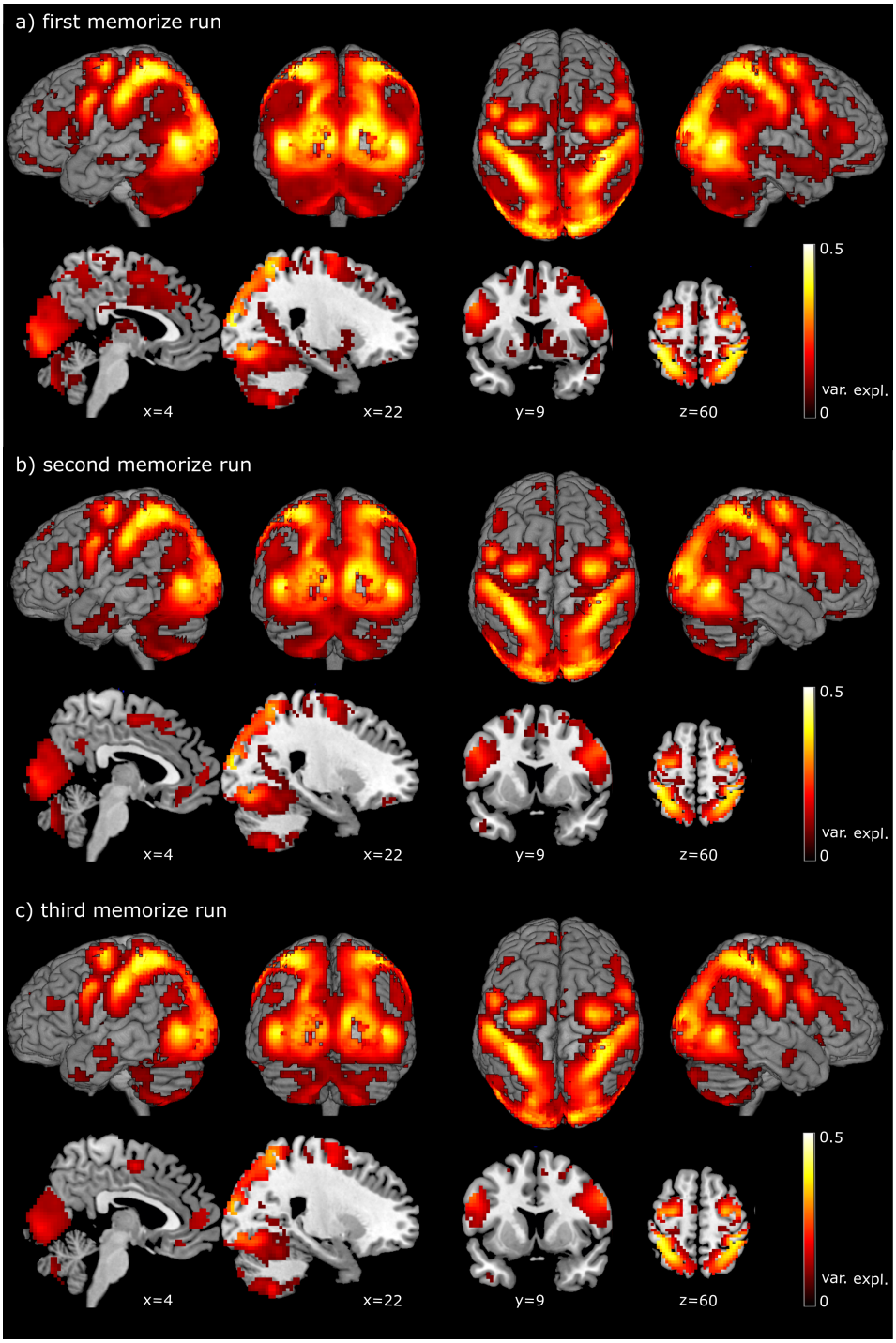


Figure 3.6: Voxels where the variance explained by the first component is significant - significant between-subjects synchrony for each session in the AOEN and beyond. q<0.001 FDR thresholded. Only clusters $\geq$ 19 displayed here.

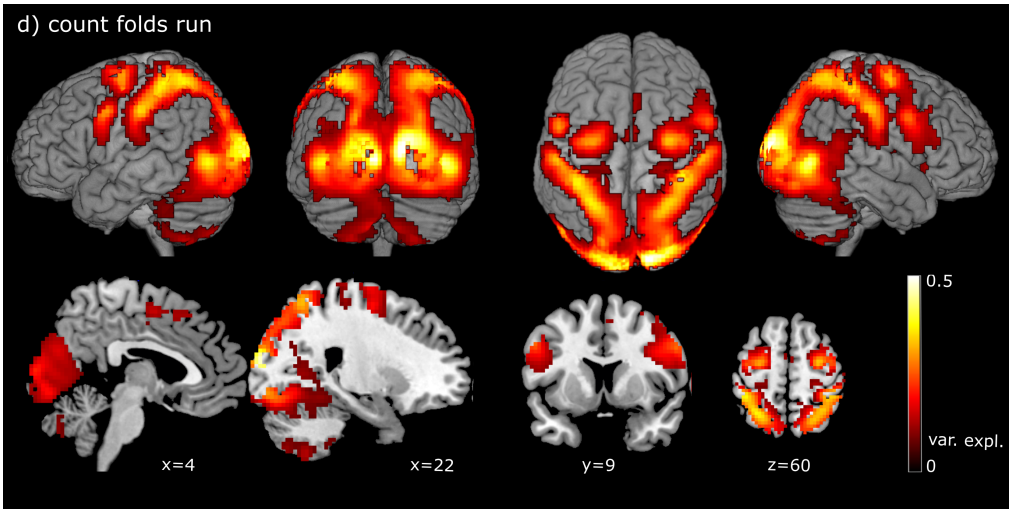


Figure 3.7: Continuation of fig. 3.6. Voxels where the variance explained by the first component is significant. q<0.001 FDR thresholded. Only clusters $\geq$ 19 displayed here.

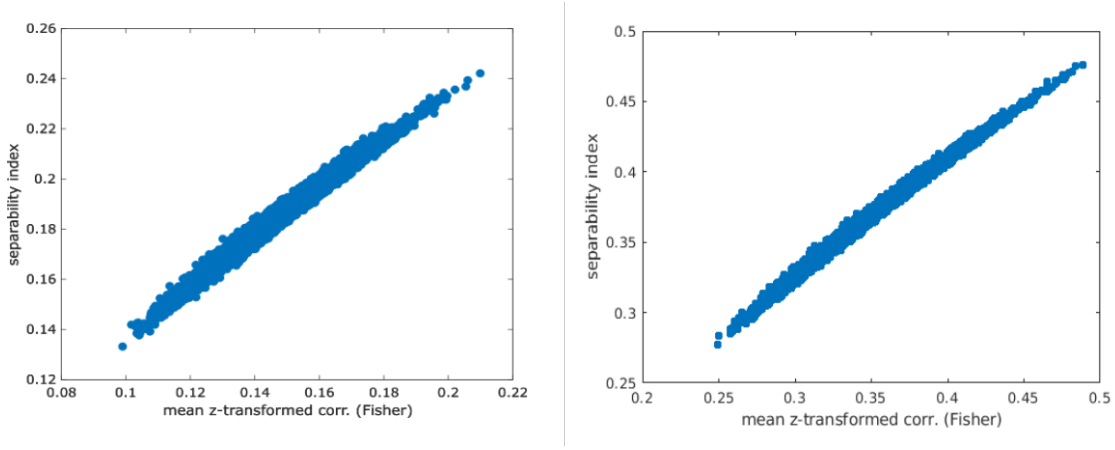


Figure 3.8: Simulated relation between separability indices and mean correlations. There is a very good fit between the mean correlations (the ISC values) and the variance explained by the first component (the depicted PCA measure). The separability index is the variance explained by the first component normalized by the total variance. Simulations for 350 data points and 28 subjects. Subject time courses are simulated as sine waves with 9 peaks with added different noise ratios. 10000 iterations per image.

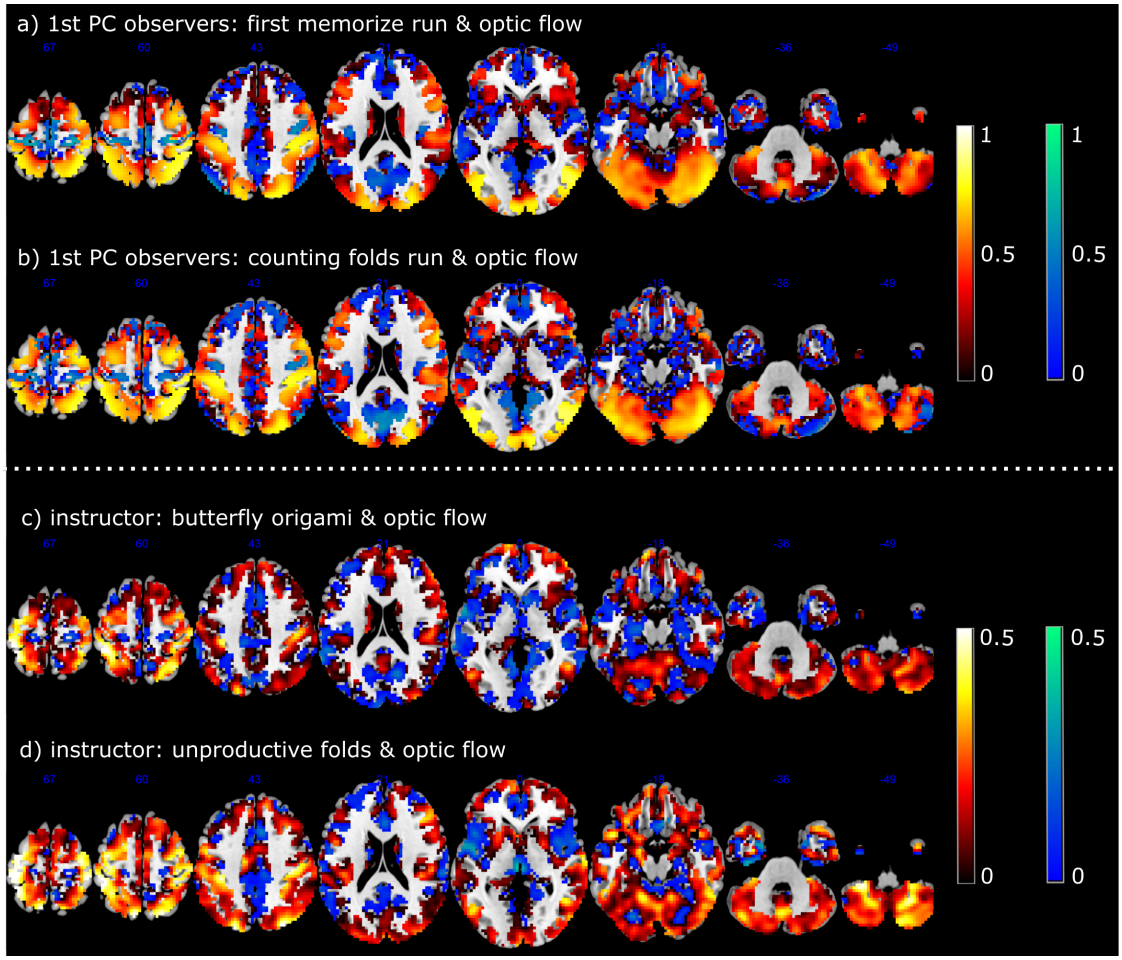


Figure 3.9: Uncorrected correlation maps. The observers as well as the instructor correlate with the optic flow in AOEN areas. a)+b) Correlation of the first principal component of the observers with the optic flow of the respective video. c)+d) correlation of the instructor data with the optic flow of the respective video.

4

General discussion

4.1 Discussion of the main findings

4.1.1 Overview

This thesis investigated the voxel-wise brain synchrony for different conditions of naturalistic action observation in the context of instructive origami videos. We researched the initial phase of learning to fold the origami and compared it to unproductive folding with no required learning. This thesis deals with two different aspects of the process: in chapter 2 we researched the relation between an instructor and the observers, and in chapter 3 the relation between the observers themselves. Since this work is based on a vast literature of brain activity during performance of actions and observation thereof, it was important to blindfold the instructor to make sure that similarities to the observers are not based on visual feedback. A post-hoc questionnaire revealed that the vast majority of the observers did not notice the blindfolding, which ensured that the naturalistic concept was maintained.

With this setup we present a paradigm where the complete initial learning stage can be done inside a constrained MRI machine and MRI imaging is possible throughout the whole learning process. The learning process can be kept as naturalistic as possible inside the constraints of the machine. Both parties, instructor and observer, can be examined using this setup. This allowed also a comparison not only across different learning tasks but also across different agents.

The thesis demonstrated new findings on the action observation network and contributed to the development of methods. The results showed that the AOEN vastly synchronizes robustly across different agents and cognitive tasks: It shows consistent synchrony irrespective of the task. Nevertheless, we found task-dependent modulations. However, the difference between cognitive tasks resided especially in non-classical AOEN areas, in particular the cerebellum and here in particular also beyond classical hand-areas. This sheds new light on the involvement of the cerebellum in action processing and learning.

On the data analysis side, this thesis transfers two data science methods to the analysis of this open naturalistic fMRI design: a bootstrapping procedure for investigating temporal lags as well as the PCA as an approach to find commonalities between subjects.

4.1.2 Synchrony of temporal activity

Synchrony of the classical AOEN across all cognitive tasks With this thesis we can show that temporal synchrony exists within the group of observers when observing different actions and performing different cognitive tasks (learn/memorize steps, monitor steps or count steps). This temporal synchrony exists between observers and instructor in the absence of visual input (our blindfolded instructor) across all different actions and different cognitive tasks. Within-observer synchrony spatially overlaps with instructor-to-observer synchrony to a great extent, compare fig. 2.4 and 3.1. Since the two figures already portray the conjunctions over all experimental conditions for the respective part of the analysis, we can conclude that in all investigated conditions in this thesis, a large proportion of the participants' brains show temporal alignment. The areas essentially cover the core structures of the Action Observation Execution System (AOEN) which lets us conclude that the AOEN synchronized robustly across all subjects for all different cognitive tasks investigated in this thesis. While an increase of activity of the AOEN during different tasks of action processing is established, we can expand this now to temporal synchrony. While this has been partly shown for pure action observation (Nummenmaa et al., 2014; Thomas et al., 2018) among subjects, we can expand this now to observational learning. Additionally, we can report a temporal coupling between instructor and observers not related to own visual feedback and independent of task.

Modulation of synchrony for different cognitive tasks especially in non-classical AOEN areas Both for the correlation to the instructor and within the observers we find parts of the parietal cortex and large parts of the cerebellum showing changes in similarity. Additionally, for the within-observer synchrony, between-task changes are located in the LOTC and partly premotor areas. The parietal cortex, LOTC or premotor cortex are established areas of the AOEN; the synchrony changes reflect different processing aspects of the tasks in the AOEN. The cerebellar synchrony changes, however, are a very new finding. They were located partly in putative cerebellar areas of the AOEN (hand sensorimotor areas), but to large extents beyond this. For the synchrony changes between tasks in the instructor-observer relation as well as between-observer relation, we find high changes within the cerebellar VI/crus 1 area. This area shows a good overlap with the spatial memory task in King et al. (2019). It also shows a good overlap with the patterns of attention, object recognition, visual working memory, and

active maintenance in the parcellation presented in the aforementioned reference. All of these described cognitive tasks are necessary for memorizing the shown origami in order to reproduce it. Consequently, synchrony changes during different tasks of action observation appear to be guided by diverse cognitive tasks leading beyond classical areas of the AOEN. Our results highlight the role of the cerebellum in action observation and learning. Note that this thesis solely uses measures of temporal synchrony, and no activation changes. Therefore it is possible that activity changes are different (for a differentiation between magnitude-based measures and correlation-based measures see appendix of Thomas et al. (2018)).
A lot of these reported cognitive tasks relate to attentiveness and cognitive control. There are some studies suggesting a relation between these and intersubject correlation (see e.g. Campbell et al., 2015 or Ki et al., 2016). However, this question needs more systematic research. Moreover, considering the sparse literature on the cerebellum, these parts remain speculative. We are presenting a likely possibility for involved cognitive functions here that nevertheless needs further investigation.

4.1.3 What areas belong to the AOEN?

This thesis shows consistent synchrony for all 'classical' AOEN areas: the motion sensitive LOTC, the object-spatial relation processing superior and inferior parietal cortex, the somatosensory cortex, the dorsal premotor cortex and the ventral premotor cortex, which includes BA44. These classical areas show a robust involvement across agents and tasks. Interestingly, in this thesis we have found effects in areas which area not that well established: the cerebellum and the primary motor cortex.

Cerebellum While the cerebellum is known to be involved in motor control, less is known for action observation. The cerebellum is a very sparsely investigated area in fMRI, especially in comparison to cortical structures (Fair, 2018). While there have been few people working in this domain of representation of cognitive tasks in the human cerebellum (e.g. Stoodley and Schmahmann, 2009; Stoodley et al., 2012), only very recently has more evidence been added. The cerebellum as a very neglected area actually encodes a vast majority of cognitive tasks (King et al., 2019) and even shows a retinotopic structure (van Es et al., 2019).
Our results of action observation fit well to the ones reported in King et al. (2019) and Abdelgabar et al. (2018). This means that the two main papers for this thesis which have performed action observation imaging with a focus on the cerebellum are both in preprint status only (meaning prior to peer-review) at the time of writing this thesis. In contrast to theses sources, we report not an increased activation, but a temporal matching of the subjects activation time courses. A temporal matching between action production and

action observation in these cerebellar areas, as we show in chapter 2, has not been shown at all before.
I therefore conclude that at least cerebellar areas VI and VIII a and b should be included in further AOEN research as we can show consistent correlations in these; the current data strongly suggests that these areas should be incorporated into the AOEN. But it will of course also be of interest to investigate the relation to other cerebellar areas like crus 1 and 2, as they have shown involvement and modulation in the tasks, and have been associated with related tasks like visuo-spatial working memory (Stoodley et al., 2012; Küper et al., 2016).
Taken together, I strongly suggest further research about the cerebellar involvement in action production, observation and learning.

Primary motor cortex When investigating the commonness of the observers we found a common principal component of the observers also in parts of the cortex which has been only associated with action execution before. For action observation, a down-regulation has been suggested (Gazzola and Keysers, 2009); the figures of Thomas et al. (2018) suggest a low intersubject correlation during action observation. Typically, no M1 involvement is reported for action observation. Nevertheless, non-human primate literature suggests that M1 contains mirror neurons, albeit some with suppression behavior (Dushanova and Donoghue, 2010; Vigneswaran et al., 2013 and Kraskov et al., 2014). Also in our case, the explained variance of the components is low. The correlations with the optic flow (putative motion proxy) are also negative, which would support the suppression characteristic. But note that it does not necessarily mean that every subject loads with the same mathematical sign onto the component, given the characteristics of the PCA analysis. Perhaps this feature helped us detect this behavior, as both positive as negative contributions to the component are valid - this would be consistent with the notion of suppression and 'normal' mirror neurons. But please note that, as always with fMRI analysis, single neuronal findings can only be one possible explanation as the fMRI signal can be generated by different properties and interplays of neurons.
Taken together, this thesis shows findings that suggest that the M1 can be involved in action observation processes. However, the explained variance of the components is low, suggesting that there are either other processes going on in M1 or that the neural contribution is so low that it is only slightly bigger than the thermal noise. Therefore I argue that this finding, albeit very interesting, is not as clear as the cerebellar findings. I also strongly encourage further research as this effect is weak and should therefore be verified by another independent experiment.

4.1.4 Reflections about behavior and performance

In this thesis I have presented a performance correlation between instructor and observers: subjects who perform better tend to show higher synchrony to the instructor in the left ventral premotor cortex during the preceding observational learning run. This classical mirror neuron area (Rizzolatti and Craighero, 2004) has been associated with 'action understanding' (Rizzolatti et al., 2002) by directly mapping observations onto own motor repertoire (the so-called 'direct-matching hypothesis', Rizzolatti et al., 2001). This is similar to the 'simulation theory' of cognition (Jacob and Jeannerod, 2005), although the extent of this theoretical framework has been questioned (Jacob and Jeannerod, 2005). Apart from these theoretical considerations, also newer studies link premotor activity modulation to learning, as elaborated in section 2.4. Also in the light of the fact that research on the brain-to-brain relation between instructor and observer in a learning setup using fMRI has -to the best knowledge- not been done yet, this finding shows a very interesting aspect of learning from an instructor.

Behavioral performance can be assessed differently. In this thesis we have addressed this by the simplest approach: counting correct steps. We give one point for each correct fold/step and half points for folds with slight faults, as especially for the 3D folds some subjects tended to give up if the fold simply wouldn't stick. It is natural to believe that the step was memorized, but just could not be executed correctly. One reason for the origami choice was to make sure subjects focus on the actual immediate action as the previous step disappears with the next folding. In comparison to, e.g., the construction of a toy model as in Frey and Gerry (2006), one cannot infer a missed step from the subsequent presented object. However, this way we cannot completely rule out that subjects have in some way memorized future steps even though they missed a step in between and even though the instruction stated 'memorize as far as you get'. On the other hand, this leads to the question to what extent it is possible for something to be memorized when it cannot be executed. When designing the experiment, I also thought about the option of placing a person next to the scanner. This person could be a possible aide: whenever the subject does not know how to proceed, she or he could give the origami to the aide and ask for the next step. However, when trying this out outside the scanner with a small pilot sample, subjects reported that it just confused them and that they couldn't continue afterwards. The chaining of motor sequences might play a role here, but this situation needs to be investigated in a complete study.

These thoughts lead to the bigger question of what actually is learned or memorized behavior. Is it only the behavior one can correctly reproduce, or is it also an already formed memory trace which cannot be translated (yet) into a motor output? This question could only be solved via a recognition test measured by correct answers or speed (see e.g. Cross et al., 2012 for a study on tying knots operating on a recognition/perceptual discrimi-

nation task). This way, no motor output is needed. It would be especially interesting to compare recognition performance to reproduction performance, and to investigate a difference in neural processing between these two already during action observation.
Also, there is another potential aspect of performance not captured in our simple measure: Subjects take more or less time to fold. Consequently, one could argue that a subject taking longer (because he or she is taking time to think or trying out folds) has memorized less. (Also, along with this, another question arises: Does this subject take longer because she or he does not remember correctly or because that person cannot translate memory into motor output?). Consequently, speed would be an interesting measure as well. However, it would be important to 'normalize' the time by a typical duration that that person takes to make a fold, to account for a person-inherent speed. It should be noted that in this setup, the subjects have not been told that they are being judged based on time (except for telling them that there is an upper boundary). From a phenomenological and descriptive point of view, some subjects seemed to fold the beginning folds during the second trial faster than during the third trial, arguably because they wanted to get to the previously missed parts as soon as possible, possibly because they were scared of forgetting due to high working memory load. Nevertheless, this would have to be analyzed in a systematic way. Taken together, duration as a behavioral measure in a naturalistic experiment needs to be handled with care, unless the subjects are instructed to also maximize speed. These thoughts also lead to another aspect: for a human spectator it seems to be easy to give a judgment about the overall performance (e.g. as I can see that a person is struggling with the motoric part or is thinking deeply or is constantly recorrecting.) Normally, these subjective ratings are avoided at all cost in order to use measurements of more objectivity, like duration or correctness of a step. Nevertheless, from a social point of view, this rating might also be a valuable parameter to give a more holistic rating, albeit possibly deviating from the objective measures. Other studies on naturalistic stimuli (e.g. Stephens et al., 2010) have engaged a group of independent raters blind to the purpose of this study. This could also offer a possible rating of behavior in a more social context.

4.2 Further research ideas

4.2.1 Expanding knowledge using this data set

This thesis investigated the similarity of single voxels over time during different conditions of action observation. This is a straight forward approach to an open design backed up by a not to be considered small number of publications based on this method. Nevertheless, as this is an open design, different aspects of relation and similarity within and between-brains could be considered.

Mapping by pattern similarity For example, in this thesis we do not investigate cortical patterns: information about differences in processes might not only be stored in the temporal behavior of one voxel, but could also be stored in patterns of neighboring voxels. We can assess this by multivariate pattern analysis (MVPA). One way of this in fMRI is to assess the representation similarity and changes thereof. This popular approach was presented by Kriegeskorte et al. (2008) and has been expanded since then. Note that here we gain information about patterns, but need to temporarily bin the information, also for computational reasons; this way we effectively lose temporal resolution. The representation similarity approach has been applied to AOEN research and shown to offer different patterns for different types of actions (see e.g. Oosterhof et al., 2010). A natural point of temporal averaging/binning might be the different steps of the origami. Consequently, a pattern similarity approach that averages across steps might offer fine-grained information about the steps. Also, here the patterns of the instructor and of the observers could be compared.
But for this study's data set, the MVPA approach offers its advantages in a different aspect: During this study, the subjects were also scanned when attempting to fold the origami. Currently, this has only been analyzed on a behavioral level, as the methods used in this thesis require a temporal alignment which is not given as the subject's foldings vary largely in time. However, when annotating the different steps and time points and averaging over a subjects' folding, we could investigate the pattern similarity during the (arguably very varying) subjects' performance. These patterns could be compared to the resulting patterns of their observation, as well as to the instructor's patterns. I hypothesize a similarity in the AOEN in all cases, even though there is an absence in temporal voxel-similarity. However, the exact locations (and potential similarities beyond) would inform us in more detail about different aspects of action processing and observation. This approach has been shown for the observation and spoken recall of a feature film (Chen et al., 2016b; Zadbood et al., 2017). It would also be interesting to relate this to the performance, as there might be a pattern difference between remembered and non-remembered movements; also, pattern similarity to the instructor might vary in this regard.

Mapping by time-warping Another approach to map the subjects' performance scans onto observation and onto the instructor would be a voxel-wise time-warping (Silbert et al., 2014). Here, detailed annotations of the performance videos would need to be made, and the timeseries needs to be warped between those computation points. However, it is not clear if this method will be feasible for the observers' performance given the very varying behavior of the participants. While Silbert et al. (2014) have shown a mapping of several speakers upon each other, these speakers were highly trained to reproduce the speech exactly like the original.

Investigating networks Connectivity differences and changes during action observation and action execution also pose an interesting approach. Similar to the logic for pattern analysis, information might be stored in changes in a subject's internal brain connectivity structure. Andric et al. (2016) have shown connectivity changes during repeated viewing of the same video, and brain network structure has been shown to change with practice for serial reaction times tasks (SRTT) (e.g. Bassett et al., 2015). By comparing the network structures during the learning conditions to the network structure of the control condition video, we could gain insights into the learning process during observation; and by comparing observation to execution we could gain insights into the network structure of the AOEN. (Note that while the AOEN is by wording a 'network' it is not defined by connectivity definitions; also typically research here focuses less on the network properties but more on common activations.) By using, e.g., independent component analysis, we could also extract common time courses of networks (similar to the common time courses of the PCA) and correlate these with parameters. By looking e.g. at the optic flow, one could research the processing on a network level.

Connecting non-matching voxels We could also explore in more detail how the instructor relates to the observers in non-matching voxels. Consequently, we could drop the voxel-wise matching and explore a connection between instructor and observers with Granger causality in Schippers et al. (2010), to see which region/seed drives what parts of the observers. Note, however, that adaption to the classical Granger causality approach may be needed as the high auto-correlative nature of especially the multiband imaging data needs to be addressed; and exact temporal relations might be blurred due to the auto-correlative nature.
Another interesting approach to combine network analysis with intersubject similarity is the so-called intersubject functional correlation presented by Simony et al. (2016). Here, a seed region or seed voxel in one person is correlated with other voxels or areas in other persons. In our case, comparing the instructor to the observers will be of interest: By taking a motor seed of the instructor, new insights on the relation of the network structure between instructor and observers could be gained. This would be especially

interesting given the findings of primary motor areas presented in this thesis. Based on our findings, I could hypothesize an anti-correlation between the somatomotor network (or, more likely, parts thereof) in instructor and observers.

4.2.2 Further research ideas for follow-ups

While this experiment yields interesting results about naturalistic learning by instruction and observation of similar stimuli, this can merely be regarded as a starting point. To broaden our knowledge, we propose several aspects for follow-ups and new experiments:

Generalize the instructor One of the main limitations of this thesis is the fact that we only have one instructor. While one instructor is naturalistic, and we can only show one video to the observers, one can still question how generalizable this one instructor is (see also section 2.4.7 for a discussion about this). As pointed out in section 2.4.7, one idea for a follow-up is to form dyads. While this offers a great option to generalize the instructor-observer coupling, it reduces the comparability between the observers and therefore would only be a follow-up for the first part of this thesis.
The second part of the thesis, though, suggests an option to involve several instructors. Silbert et al. (2014) have mapped three highly-trained speakers upon each other by a time-warp algorithm using the audio envelope of the spoken signal. If we wanted to warp performing subjects onto each other, the first intuitive approach to take would be to annotate the videos manually to match similar steps and then dilate and stretch the voxel time courses in between the matching points. This approach of annotation would be tedious and have the risk of being subjective. However, section 3.4.2 of the second manuscript would offer a possible method of generalization: by mapping the optic flow of the resulting production videos (all taken from the same perspective) upon each other, we could possibly warp the voxel-time courses upon each other. (If we want to generalize to the whole brain, we would have to make the assumption that the other unrelated processes are also stretchable in time in a similar way.). This would offer a possible way of investigating several instructors even though performance varies naturally. This way we could show a reliability in the presented instructor.
Additionally, one should also test the instructor(s) under different conditions. Especially the case of visual feedback deserves closer attention: In our case, the instructor is blindfolded to show the direct action-observation matching; it is vital here to rule out visual feedback. Nevertheless, there may be a difference between performing an action with, or without visual feedback. Literature from non-human primates suggests there are mirror neurons that show different reactions based on the presence or absence of visual feedback (Maranesi et al., 2015). This could indicate a different activation pattern in humans based on visual feedback. But even without inquiring mirror neurons

visual feedback may play a role and should be investigated especially in comparison to the blindfold. For example, it is not unreasonable to assume that in the blindfold case haptic/somatosensory feedback may play a bigger role during actions. Especially the literature on a putative human mirror neuron system required the performing of actions in darkness (Turella et al., 2009) (see also e.g. Gazzola and Keysers, 2009).
Note that this also shows a classical conflict in performing naturalistic experiments - the tension between control and becoming less naturalistic. In our case, blindfolding the instructor was possible, in contrast to the study on spoken language of Stephens et al. (2010). Here, they had to accept the possibility of the confound that the similarity arises from the speaker listening to her/his own speech. On the other hand, even though we have made sure that the naturalistic concept was kept by the choice of origami and practice (and checked by a follow-up questionnaire), we cannot rule out the possibility that due to the blindfold we are comparing unnatural instructor activity. Also for this reason, I strongly suggest comparing the presence and absence of visual feedback for the instructor.

Generalize task For this experiment the subjects effectively only learned one origami. To generalize the results, different origami would be helpful. Especially a longer duration would be interesting: Do we have a reduction of synchrony to the instructor with longer duration due to e.g. attention/fatigue effects? Here, we could also employ a measurement of variable synchrony like a sliding-window correlation (see e.g. Nguyen et al., 2017).
Needless to say, when investigating the initial stage of learning, it is important not only to generalize to different paper figures with different complexity, but also to different tasks. While a lot of research has been done on sequence learning tasks involving keyboards, it is important to consider other tasks, as a lot of naturalistic actions also involve a visuo-spatial and a motor-adaptive component which are not much present in keyboard presses. So far, knot-tying (Tracy et al., 2003; Cross et al., 2012, 2017) and the construction of a toy model (Frey and Gerry, 2006) have been investigated, but under partly different aspects and partly without immediate reproduction. Certainly, more research should be done here. Even more importantly, these tasks should be compared.

Behavioral relation and attentiveness To reproduce the paper foldings (or other parts of a sequential progress) sustained attentiveness to the demonstration (the video) is necessary; as in general for the beginning stage of observational learning where information needs to be absorbed (Bandura, 1971). However, for the complete process of memorization and reproduction, attentiveness is not sufficient and other cognitive processes play a role. Consequently, not only the reproduction but also the attentiveness may be of interest to the similarity within observers or to the instructor. Measuring attentiveness in a naturalistic setup is tricky, though. One option is a post-hoc self-report

as in Nummenmaa et al. (2012) but self-reports need to be handled with caution. Another option would be a dual MRI-EEG setup to measure alpha power as an established marker for attentiveness. However, Ki et al. (2016) have shown that in EEG, ISC itself may be an even better marker for attentiveness than alpha power. Consequently, it is very important to also address this question.

4.3 Relevance of this thesis

4.3.1 Direct contributions

This thesis provides the following contributions to the scientific community:

Methodology First of all, this thesis provides methodological contributions: In section 2.2.6 we introduce a method to assess lags to fMRI, to determine whether one participant's brain activity advances that of another participant. fMRI data shows a high autocorrelative structure in general and with the faster imaging of the newer multiband imaging technique this problem is enhanced. We therefore suggest a bootstrap routine to assess lags which does not assume an a-priori distribution. Traditionally, lags are determined via the maximum of the crosscorrelation of two time courses (see e.g. Kreuz et al., 2007 and for fMRI Zadbood et al., 2017). However, in the case of noisy data this just shows a possibility that there is a lag. Statistical testing is needed to assure that the lag is significant. We present a method for this.
Also we present a PCA approach to assess the commonness of the data, instead of calculating pairwise correlations. With this method we can extract common time courses and determine the significance and strength of the relationship. A related approach has been suggested by Hanson et al. (2009). Given the popularity of the PCA approach in other domains (Abdi and Williams, 2010b) and also in today's data science trend, we would like to show the possibilities of this approach, while keeping the relation to the previous literature, as the variance explained by the first component is highly related to the mean average correlation of the correlation matrix, see section 1.1.4.

Action-observation-execution research One main finding of this thesis is the shown overall temporal synchrony between the participants - between the observers as well as between the observers and the instructor. This synchrony exists for the different stages of initial learning as well as pure action identification. The synchronized areas are the same ones as previously shown to be involved in action-observation-execution tasks for classical magnitude-based research - plus the previously seldom reported cerebellar hand areas. We therefore demonstrate that the participants synchronize temporally and that the AOEN synchronization is robust across the different cognitive tasks. This indicates

a similar processing in AOEN areas during action without observation (blindfolded instructor) and observation without action - the latter with the task to identify (count folds) or to memorize to reproduce or to monitor just reproduced actions. Our research suggests that the AOEN is consequently a robust 'always-there' system with similar processing across agents.
Moreover, we can show fine-grained changes in synchrony for different cognitive aspects of the tasks. As an overall main effect with only small exception areas, the subjects synchronize the most in the first learning run, meaning they are the most similar to each other as well as to the instructor in this run. Part of the synchrony reductions in comparison to the other sessions is located in parietal areas or the LOTC. These are classical AOEN areas previously associated with motion or spatial processing. Consequently, a possible explanation is that less (similar) visuo-spatial processing is utilized by the subjects. The most novel finding of this thesis is that an important part of the synchrony reductions -in the case of the instructor-observer synchrony the greatest part of synchrony reductions by far- is located in in cerebellar areas. The cerebellum has previously been a neglected area in fMRI research; this thesis demonstrates that it plays a vital and non-negligible role in action observation and learning. Based on the available sparse literature, we conclude a higher cerebellar synchrony of visuo-spatial working memory, cognitive control or multisensory integration. Our results go well beyond the classical interpretation of the cerebellum as a pure 'motor control' organ.
Additionally, our data-driven approaches have unraveled other fine-grained aspects of AOEN function: our analysis shows that the observers' activity anticipates the instructor's activity in the aIPS, as a potential expression of immediate goal anticipation of the observers. Moreover, analysis of the optic flow pattern revealed a strong encoding of this information in the AOEN, potentially an indicator of hand motion encoding in the complete system.
Lastly, this work is also relevant as, to my knowledge, no object-forming task except for tying knots has been investigated using fMRI before - this is especially interesting as the task bridges sequence learning (like pressing keyboard sequences) and visuo-spatial learning of producing shapes of specific spatial orientation.

Social neuroscience Finally, this thesis informs social neuroscience as it investigates naturalistic human-to-human interaction with both parties being scanned in the MRI. Interaction research especially using fMRI is far from common. Consequently, we are the first to show a coupling or synchrony between instructor and observers in a learning setting - but also for mere observation. Importantly, this cannot be explained by common visual feedback as the instructor received none. Instead, it has to rely on more specific cognitive processes of synchronizing the observers' brain activity to the instructor's, who has been scanned first. The data has also shown that a stronger synchronization to

the instructor can be associated with a better subsequent performance in a part of the premotor cortex - a potential benefit of this social interaction.

4.3.2 Outlook

Even though this thesis investigated an every-day task, it is to be considered basic research. Basic research in cognitive neuroscience typically informs two neighboring disciplines: medicine and psychology/cognitive science. By determining the neural involvement in cognitive processes, we can inform new psychological hypotheses e.g. about learning. This thesis offers merely a starting point for a learning sciences contribution: more research needs to be done on different stages of learning by observing instructive videos. While doing so has been shown to be effective (Michas and Berry, 2000; Wong et al., 2009; Marcus et al., 2013) and is a common practice in today's digital world, little neuroscientific research has been done in this field. In the future it will be especially relevant to pay attention to different subject groups. Young and healthy subjects are always the cohort of first choice for all kinds of psychological or neuroscientific experiments as typically university students are the easiest subjects to recruit. Nevertheless, especially for older people, it will be important to investigate cognitive and neuronal processes during video learning and compare the results to a young cohort. Understanding different age groups and bridging potential gaps between them will be also of societal importance. If we consider the digital transformation processes, learning and re-learning by observing instructive videos (in the bigger context of 'e-learning') will play a great role in the future (Dittler et al., 2017). Understanding differences between people could inform e.g. personalized learning options.

Another positive aspect of the digital development is the increasing computational power. Along with this continuing increase, many new data-driven approaches have been created in cognitive neuroscience (see also chapter 4.2 for examples of other data-driven approaches). These approaches often come with a heavy computational load which poses high demands not only on working memory but also requires potentially long computational time. Consequently, the practical feasibility has only been reached recently and is still improving. On the positive side, this gives the option of opening up the experiment designs to less controlled but more naturalistic stimuli. In this sense, we can use instructional videos for experiments without controlling for e.g. amount and types of movement. This way we can use everyday stimuli to inform us about everyday cognitive processes.

As stated above, basic research in cognitive neuroscience can also inform clinical research: Data-driven designs in general offer the possibility to perform less-controlled experiments. This can be advantageous for clinical research as these designs tend to put less cognitive demand on the subject or patient. An extreme case of an uncontrolled experiment with

literally no cognitive demand popular in clinical research would be resting state fMRI. But also tasks like movie-watching would be similarly easy to establish. Following these lines, naturalistic experiments offer an advantage for clinical research as the tasks are closer to the actualities of the subjects' or patients' lives and for this reason often require less cognitive ability for task understanding or performance. Consequently, it is also very likely that the compliance will be higher.

Naturalistic experimental designs are of importance especially for psychiatry. Here, the advantage of data-driven designs is in line with the advantages these designs have for social neuroscience: social stimuli are in themselves very complex, meaning that encapsulating stimuli parts for exact hypotheses is often impossible or not advisable (Adolphs et al., 2016). If we, for example, depict a human interacting with us, several aspects potentially play a role: appearance, posture, emotion, expression, wording, motion etc. Interaction itself can arise differently from the combination and interplay of these factors than for each factor individually; we cannot assume that cognition is a summation process. An alteration of these cognitive processes is characteristic for many psychiatric conditions, see e.g. the diagnostic criteria of the DSM-5 (American Psychiatric Association et al., 2013). Consequently, a strong link between psychiatric diseases and impairment of social cognition has been stated by Schilbach (2016). The author explicitly requests a two-person approach to neuropsychiatry. Along these lines, data-driven open designs have informed psychiatry with e.g. differences in video processing and judgment associated with paranoia (Finn et al., 2018) or low intersubject correlation during movie watching in autism (Hasson et al., 2009).

For autism research, observational learning may be of particular interest: Autism has been previously suggested to be associated with altered function of the mirror neuron system (see e.g. Rizzolatti et al., 2009; Rizzolatti and Sinigaglia, 2016 for reviews), also sometimes referred to as the 'broken mirror theory' (Ramachandran and Oberman, 2006). This theory was especially popular during the main phase of explicit mirror neuron research in humans from 2000 to the first half of the 2010s. However, contradicting results about e.g. impaired imitation capability have been found, as e.g. (Hamilton, 2013) pointed out. Another hypothesis for autistic dysfunction is an impaired mentalizing ability and/or chaining of actions (Hamilton, 2013). Investigating naturalistic sequence learning as a task could offer a new approach here. Especially when varying the social context (pictographic instructions, animations, real hand actions as in Marcus et al. (2013), or presenting a real instructor). The direct comparison of the brain activity to the instructor brain activity when following a setup similar to the one presented in this thesis could also be interesting. Importantly, an open data-driven design should investigate the whole brain, not only in order to look beyond mirror neurons but also to look at previously neglected brain areas:

Stoodley et al. (2016) and Schmahmann (2019) distinguish two mayor types of deficits after cerebellar lesions: the cerebellar motor syndrome resulting in impaired motor abilities, the cerebellar cognitive affective syndrome (CCAS), as well as combinations of these two. The cerebellar CCAS manifests in 'autism spectrum and psychosis spectrum disorders, and disorders of emotional control, attentional control and social skill set' (Schmahmann, 2019). While lesions leading to the cerebellar motor syndrome (without CCAS) are found in the anterior cerebellum, lesions leading to CCAS (without motor syndrome) are found in crus 1 to lobule IX. Since in this thesis we have found changes in cerebellar similarity in the latter areas for different types of action observation, this poses an interesting question about whether an impairment here could be associated with autism, since a lesion in these areas could. While autism has been suggested to be associated with a dysfunctional MNS/AOEN, this research was strongly focused on only cortical involvements. I therefore strongly suggest that the cerebellar involvement in different action observation tasks not only in healthy subjects, but also in a clinical context should be investigated.

4.3.3 Conclusion

Taken together, naturalistic imaging and in particular naturalistic observational learning are of interest in basic cognitive neuroscience, for learning sciences, and for clinical research. Today's increasing computational power makes analyzing these data-driven designs more and more feasible. In this thesis, I have designed a paradigm that enables the recording and analysis of the brain activity of a complete observational learning process. With this setup it was possible to demonstrate new aspects of the AOEN. As a main finding, the results show a robust synchrony across different cognitive tasks and agents on the one hand, and synchrony changes across tasks in areas not previously associated with the AOEN like cerebellar crus 1 and 2 on the other hand. Additional data-driven approaches have unraveled other fine-grained aspects of the AOEN.
While strongly controlled cognitive experiments are certainly of great importance, I hope that naturalistic designs will play an important role as well in the future.

Bibliography

Abdelgabar, A. R., Suttrup, J., Broersen, R., Bhandari, R., Picard, S., Keysers, C., De Zeeuw, C. I., and Gazzola, V. (2018). Action perception recruits the cerebellum and is impaired in spinocerebellar ataxia patients. *bioRxiv*.

Abdi, H. and Williams, L. J. (2010a). Jackknife.

Abdi, H. and Williams, L. J. (2010b). Principal component analysis: Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459.

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., and Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience*, 37(9):1458–1469.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour: Cognitive neuroscience. *Nature Reviews Neuroscience*, 4(3):165–178.

Adolphs, R., Nummenmaa, L., Todorov, A., and Haxby, J. V. (2016). Data-driven approaches in the investigation of social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693):20150367.

American Psychiatric Association et al. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., and Haynes, J.-D. (2011). Flow of affective information between communicating brains. *NeuroImage*, 54(1):439–446.

Andric, M., Goldin-Meadow, S., Small, S. L., and Hasson, U. (2016). Repeated movie viewings produce similar local activity patterns but different network configurations. *NeuroImage*, 142:613–627.

Apšvalka, D., Cross, E. S., and Ramsey, R. (2018). Observing action sequences elicits sequence-specific neural representations in frontoparietal brain regions. *The Journal of Neuroscience*.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113.

Bandura, A. (1971). Social Learning Theory. General Learning Press, New York.

Bandura, A. (2004). Observational Learning. In Byrne, J. H., editor, *Learning and Memory*, pages 482–484. Macmillan Reference USA, New York, 2nd ed. edition.

Bartels, A., Zeki, S., and Logothetis, N. K. (2008). Natural Vision Reveals Regional Specialization to Local Motion and to Contrast-Invariant, Global Flow in the Human Brain. *Cerebral Cortex*, 18(3):705–717.

Bassett, D. S., Yang, M., Wymbs, N. F., and Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nature Neuroscience*, 18(5):744–751.

Berg, A., McMurry, T. L., and Politis, D. N. (2010). Subsampling p-values. *Statistics & probability letters*, 80(17-18):1358–1364.

Bhat, A. N., Hoffman, M. D., Trost, S. L., Culotta, M. L., Eilbott, J., Tsuzuki, D., and Pelphrey, K. A. (2017). Cortical Activation during Action Observation, Action Execution, and Interpersonal Synchrony in Adults: A functional Near-Infrared Spectroscopy (fNIRS) Study. *Frontiers in Human Neuroscience*, 11.

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V. D., Schmahl, C., Demanuele, C., Tost, H., Kirsch, P., and Meyer-Lindenberg, A. (2015). Information flow between interacting human brains: Identification, validation, and relationship to social expertise. *Proceedings of the National Academy of Sciences*, 112(16):5207–5212.

Binkofski, F., Amunts, K., Stephan, K. M., Posse, S., Schormann, T., Freund, H.-J., Zilles, K., and Seitz, R. J. (2000). Broca's region subserves imagery of motion: A combined cytoarchitectonic and fMRI study. *Human Brain Mapping*, 11(4):273–285.

Binkofski, F. and Buccino, G. (2006). The role of ventral premotor cortex in action execution and action understanding. *Journal of Physiology-Paris*, 99(4-6):396–405.

Brandt, T., DeutschläNder, A., Glasauer, S., Nolte, A., BrüCkmann, H., Dieterich, M., and Stephan, T. (2005). Expectation of Sensory Stimulation Modulates Brain Activation during Visual Motion Stimulation. *Annals of the New York Academy of Sciences*, 1039(1):325–336.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., and Rizzolatti, G. (2004). Neural Circuits Underlying Imitation Learning of Hand Actions: An Event-Related fMRI Study. *Neuron*, 42(2):323–334.

Calvo-Merino, B., Glaser, D., Grèzes, J., Passingham, R., and Haggard, P. (2005). Action Observation and Acquired Motor Skills: An fMRI Study with Expert Dancers. *Cerebral Cortex*, 15(8):1243–1249.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. (2006). Seeing or Doing? Influence of Visual and Motor Familiarity in Action Observation. *Current Biology*, 16(19):1905–1910.

Camilleri, J., Müller, V., Fox, P., Laird, A., Hoffstaedter, F., Kalenscher, T., and Eickhoff, S. (2018). Definition and characterization of an extended multiple-demand network. *NeuroImage*, 165:138–147.

Campbell, K. L., Shafto, M. A., Wright, P., Tsvetanov, K. A., Geerligs, L., Cusack, R., and Tyler, L. K. (2015). Idiosyncratic responding during movie-watching predicted by age differences in attentional control. *Neurobiology of Aging*, 36(11):3045–3055.

Carter, R. M. and Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends in Cognitive Sciences*, 17(7):328–336.

Casiraghi, L., Alahmadi, A. A. S., Monteverdi, A., Palesi, F., Castellazzi, G., Savini, G., Friston, K., Gandini Wheeler-Kingshott, C. A. M., and D'Angelo, E. (2019). I See Your Effort: Force-Related BOLD Effects in an Extended Action Execution–Observation Network Involving the Cerebellum. *Cerebral Cortex*, pages bhy322–bhy322.

Caspers, S., Zilles, K., Laird, A. R., and Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, 50(3):1148–1167.

Cerri, G., Cabinio, M., Blasi, V., Borroni, P., Iadanza, A., Fava, E., Fornia, L., Ferpozzi, V., Riva, M., Casarotti, A., Martinelli Boneschi, F., Falini, A., and Bello, L. (2015). The mirror neuron system and the strange case of Broca's area: Broca is Not Part of Mirror Neuron System. *Human Brain Mapping*, 36(3):1010–1027.

Chen, G., Shin, Y.-W., Taylor, P. A., Glen, D. R., Reynolds, R. C., Israel, R. B., and Cox, R. W. (2016a). Untangling the relatedness among correlations, part I: Nonparametric approaches to inter-subject correlation analysis at the group level. *NeuroImage*, 142:248–259.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., and Hasson, U. (2016b). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20(1):115–125.

Cohen, A.-L., Shavalian, E., and Rube, M. (2015). The Power of the Picture: How Narrative Film Captures Attention and Disrupts Goal Pursuit. *PloS one*, 10(12):e0144493.

Cook, R., Bird, G., Catmur, C., Press, C., and Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences*, 37(02):177–192.

Cross, E. S., Cohen, N. R., Hamilton, A. F. d. C., Ramsey, R., Wolford, G., and Grafton, S. T. (2012). Physical experience leads to enhanced object perception in parietal cortex: Insights from knot tying. *Neuropsychologia*, 50(14):3207–3217.

Cross, E. S., Hamilton, A. F. d. C., Cohen, N. R., and Grafton, S. T. (2017). Learning to tie the knot: The acquisition of functional object representations by physical and observational experience. *PloS one*, 12(10):e0185044.

Cross, E. S., Hamilton, A. F. d. C., and Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, 31(3):1257–1267.

Cross, E. S., Kraemer, D. J., Hamilton, A. F. d. C., Kelley, W. M., and Grafton, S. T. (2009). Sensitivity of the Action Observation Network to Physical and Observational Learning. *Cerebral Cortex*, 19(2):315–326.

Dai, R., Liu, R., Liu, T., Zhang, Z., Xiao, X., Sun, P., Yu, X., Wang, D., and Zhu, C. (2018). Holistic cognitive and neural processes: a fNIRS-hyperscanning study on interpersonal sensorimotor synchronization. *Social Cognitive and Affective Neuroscience*, 13(11):1141–1154.

Dayan, E., Barliya, A., de Gelder, B., Hendler, T., Malach, R., and Flash, T. (2018). Motion cues modulate responses to emotion in movies. *Scientific Reports*, 8(1).

Dayan, E. and Cohen, L. G. (2011). Neuroplasticity Subserving Motor Skill Learning. *Neuron*, 72(3):443–454.

De Bin, R., Janitza, S., Sauerbrei, W., and Boulesteix, A.-L. (2016). Subsampling versus bootstrapping in resampling-based model selection for multivariable regression. *Biometrics*, 72(1):272–280.

Depireux, D. A., Simon, J. Z., Klein, D. J., and Shamma, S. A. (2001). Spectro-Temporal Response Field Characterization With Dynamic Ripples in Ferret Primary Auditory Cortex. *Journal of Neurophysiology*, 85(3):1220–1234.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, 91(1):176–180.

Dinstein, I., Hasson, U., Rubin, N., and Heeger, D. J. (2007). Brain Areas Selective for Both Observed and Executed Movements. *Journal of Neurophysiology*, 98(3):1415–1427.

Dittler, U., Hofhues, S., Hübner, S., Kreidl, C., Müller, J., Rowbotham, M., Riesenbeck, W., Schwertel, U., Schumacher, G., and Wode, B. (2017). *E-Learning 4.0: Mobile Learning, Lernen mit Smart Devices und Lernen in sozialen Netzwerken*. De Gruyter.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4):172–179.

Dushanova, J. and Donoghue, J. (2010). Neurons in primary motor cortex engaged during action observation. *European Journal of Neuroscience*, 31(2):386–398.

Errante, A. and Fogassi, L. (2019). Parieto-frontal mechanisms underlying observation of complex hand-object manipulation. *Scientific Reports*, 9(1):348.

Fair, D. A. (2018). The Big Reveal: Precision Mapping Shines a Gigantic Floodlight on the Cerebellum. *Neuron*, 100(4):773–776.

Fedorenko, E., Duncan, J., and Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, 110(41):16616–16621.

Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A., and Constable, R. T. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nature Communications*, 9(1).

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science*, 308(5722):662.

Frey, S. H. and Gerry, V. E. (2006). Modulation of Neural Activity during Observational Learning of Actions and Their Sequential Orders. *Journal of Neuroscience*, 26(51):13194–13201.

Friedman, S. and Weisberg, H. F. (1981). Interpreting the First Eigenvalue of a Correlation Matrix. *Educational and Psychological Measurement*, 41(1):11–21.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456):815–836.

Frith, C. D. and Frith, U. (2007). Social Cognition in Humans. *Current Biology*, 17(16):R724–R732.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2):593–609.

Gardner, T., Aglinskas, A., and Cross, E. S. (2017). Using guitar learning to probe the Action Observation Network's response to visuomotor familiarity. *NeuroImage*, 156:174–189.

Gardner, T., Goulden, N., and Cross, E. S. (2015). Dynamic Modulation of the Action Observation Network by Movement Familiarity. *Journal of Neuroscience*, 35(4):1561–1572.

Gazzola, V. and Keysers, C. (2009). The Observation and Execution of Actions Share Motor and Somatosensory Voxels in all Tested Subjects: Single-Subject Analyses of Unsmoothed fMRI Data. *Cerebral Cortex*, 19(6):1239–1255.

Golland, Y., Bentin, S., Gelbard, H., Benjamini, Y., Heller, R., Nir, Y., Hasson, U., and Malach, R. (2007). Extrinsic and Intrinsic Systems in the Posterior Cortex of the Human Brain Revealed during Natural Sensory Stimulation. *Cerebral Cortex*, 17(4):766–777.

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., and Dosenbach, N. U. (2017). Precision Functional Mapping of Individual Human Brains. *Neuron*, 95(4):791–807.e7.

Grafton, S. T. (2009). Embodied Cognition and the Simulation of Action to Understand Others. *Annals of the New York Academy of Sciences*, 1156(1):97–117.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10(1):14–23.

Hamilton, A. F. d. C. (2013). Reflecting on the mirror neuron system in autism: A systematic review of current theories. *Developmental Cognitive Neuroscience*, 3:91–105.

Hamilton, A. F. d. C. and Grafton, S. T. (2006). Goal Representation in Human Anterior Intraparietal Sulcus. *The Journal of Neuroscience*, 26(4):1133.

Hanson, S. J., Gagliardi, A. D., and Hanson, C. (2009). Solving the brain synchrony eigenvalue problem: conservation of temporal dynamics (fMRI) over subjects doing the same task. *Journal of Computational Neuroscience*, 27(1):103–114.

Hardwick, R. M., Caspers, S., Eickhoff, S. B., and Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience & Biobehavioral Reviews*, 94:31–44.

Hardwick, R. M., Rottschy, C., Miall, R. C., and Eickhoff, S. B. (2013). A quantitative meta-analysis and review of motor learning in the human brain. *NeuroImage*, 67:283–297.

Hasson, U., Avidan, G., Gelbard, H., Vallines, I., Harel, M., Minshew, N., and Behrmann, M. (2009). Shared and idiosyncratic cortical activation patterns in autism revealed under continuous real-life viewing conditions. *Autism Research*, 2(4):220–231.

Hasson, U. and Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693):20150366.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16(2):114–121.

Hasson, U. and Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, 62(2):1272–1278.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D. J. (2008a). Neurocinematics: The neuroscience of film. *Projections*, 2(1):1–26.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, 303(5664):1634–1640.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008b). A Hierarchy of Temporal Receptive Windows in Human Cortex. *Journal of Neuroscience*, 28(10):2539–2550.

Herbec, A., Kauppi, J.-P., Jola, C., Tohka, J., and Pollick, F. E. (2015). Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance. *Cortex*, 71:341–348.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of cognitive neuroscience*, 21(7):1229–1243.

Hickok, G. (2013). Do mirror neurons subserve action understanding? *Neuroscience Letters*, 540:56–58.

Higuchi, S., Holle, H., Roberts, N., Eickhoff, S., and Vogt, S. (2012). Imitation and observational learning of hand actions: Prefrontal involvement and connectivity. *NeuroImage*, 59(2):1668–1683.

Holper, L., Scholkmann, F., and Wolf, M. (2012). Between-brain connectivity during imitation measured by fNIRS. *NeuroImage*, 63(1):212–222.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., and Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286(5449):2526–2528.

Jacob, P. and Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in Cognitive Sciences*, 9(1):21–25.

Kantak, S. S., Stinear, J. W., Buch, E. R., and Cohen, L. G. (2012). Rewiring the brain: potential role of the premotor cortex in motor control, learning, and recovery of function following brain injury. *Neurorehabilitation and neural repair*, 26(3):282–292.

Kauppi, J.-P., Pajula, J., and Tohka, J. (2014). A versatile software package for inter-subject correlation based analyses of fMRI. *Frontiers in Neuroinformatics*, 8.

Keller, G. B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. *Neuron*, 74(5):809–815.

Keysers, C. (2009). Mirror neurons. *Current Biology*, 19(21):R970–R971.

Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *Journal of Neuroscience*, 36(10):3092–3101.

Kilner, J. and Lemon, R. (2013). What We Know Currently about Mirror Neurons. *Current Biology*, 23(23):R1057–R1062.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, 15(8):352–357.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., and Frith, C. D. (2009). Evidence of Mirror Neurons in Human Inferior Frontal Gyrus. *The Journal of Neuroscience*, 29(32):10153–10159.

King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B., and Diedrichsen, J. (2019). A Multi-Domain Task Battery Reveals Functional Boundaries in the Human Cerebellum. *bioRxiv*, page 423509.

Kirsch, L. P. and Cross, E. S. (2015). Additive Routes to Action Learning: Layering Experience Shapes Engagement of the Action Observation Network. *Cerebral Cortex*, 25(12):4799–4811.

Koike, T., Sumiya, M., Nakagawa, E., Okazaki, S., and Sadato, N. (2019). What Makes Eye Contact Special? Neural Substrates of On-Line Mutual Eye-Gaze: A Hyperscanning fMRI Study. *eneuro*, 6(1):ENEURO.0284–18.2019.

Koike, T., Tanabe, H. C., Okazaki, S., Nakagawa, E., Sasaki, A. T., Shimada, K., Sugawara, S. K., Takahashi, H. K., Yoshihara, K., Bosch-Bayard, J., and Sadato, N. (2016). Neural substrates of shared attention as social memory: A hyperscanning functional magnetic resonance imaging study. *NeuroImage*, 125:401–412.

Kok, E., De Bruin, A. B., Geel, K. v., Gegenfurtner, A., Heyligers, I., and Sorger, B. (2018). The Neural Implementation of Surgical Expertise Within the Mirror-Neuron System: An fMRI Study. *Frontiers in Human Neuroscience*, 12.

Koski, L., Iacoboni, M., Dubeau, M.-C., Woods, R. P., and Mazziotta, J. C. (2003). Modulation of Cortical Activity During Different Imitative Behaviors. *Journal of Neurophysiology*, 89(1):460–471.

Küper, M., Kaschani, P., Thürling, M., Stefanescu, M. R., Burciu, R. G., Göricke, S., Maderwald, S., Ladd, M. E., Hautzel, H., and Timmann, D. (2016). Cerebellar fMRI Activation Increases with Increasing Working Memory Demands. *The Cerebellum*, 15(3):322–335.

Kraskov, A., Dancause, N., Quallo, M. M., Shepherd, S., and Lemon, R. N. (2009). Corticospinal Neurons in Macaque Ventral Premotor Cortex with Mirror Properties: A Potential Mechanism for Action Suppression? *Neuron*, 64(6):922–930.

Kraskov, A., Philipp, R., Waldert, S., Vigneswaran, G., Quallo, M. M., and Lemon, R. N. (2014). Corticospinal mirror neurons. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1644):20130174–20130174.

Kreuz, T., Mormann, F., Andrzejak, R. G., Kraskov, A., Lehnertz, K., and Grassberger, P. (2007). Measuring synchronization in coupled model systems: A comparison of different approaches. *Physica D: Nonlinear Phenomena*, 225(1):29–42.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4.

Lago-Rodríguez, A., Cheeran, B., Koch, G., Hortobagy, T., and Fernandez-del Olmo, M. (2014). The role of mirror neurons in observational motor learning: an integrative review. *European Journal of Human Movement*, 32:82–103.

Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., and Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, 100:316–324.

Lee, R. F., Dai, W., and Jones, J. (2012). Decoupled circular-polarized dual-head volume coil pair for studying two interacting human brains with dyadic fMRI. *Magnetic Resonance in Medicine*, 68(4):1087–1096.

Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *Journal of Neuroscience*, 31(8):2906–2915.

Lingnau, A. and Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, 19(5):268–277.

Mainieri, A. G., Heim, S., Straube, B., Binkofski, F., and Kircher, T. (2013). Differential role of the Mentalizing and the Mirror Neuron system in the imitation of communicative gestures. *NeuroImage*, 81:294–305.

Maranesi, M., Livi, A., and Bonini, L. (2015). Processing of Own Hand Visual Feedback during Object Grasping in Ventral Premotor Mirror Neurons. *Journal of Neuroscience*, 35(34):11824–11829.

Marcus, N., Cleary, B., Wong, A., and Ayres, P. (2013). Should hand actions be observed when learning hand motor skills from instructional animations? *Computers in Human Behavior*, 29(6):2172–2178.

Marek, S., Siegel, J. S., Gordon, E. M., Raut, R. V., Gratton, C., Newbold, D. J., Ortega, M., Laumann, T. O., Adeyemo, B., Miller, D. B., Zheng, A., Lopez, K. C., Berg, J. J., Coalson, R. S., Nguyen, A. L., Dierker, D., Van, A. N., Hoyt, C. R., McDermott, K. B., Norris, S. A., Shimony, J. S., Snyder, A. Z., Nelson, S. M., Barch, D. M., Schlaggar, B. L., Raichle, M. E., Petersen, S. E., Greene, D. J., and Dosenbach, N. U. (2018). Spatial and Temporal Organization of the Individual Human Cerebellum. *Neuron*, 100(4):977–993.e7.

Merabet, L. B., Swisher, J. D., McMains, S. A., Halko, M. A., Amedi, A., Pascual-Leone, A., and Somers, D. C. (2007). Combined Activation and Deactivation of Visual Cortex During Tactile Sensory Processing. *Journal of Neurophysiology*, 97(2):1633–1641.

Michas, I. C. and Berry, D. C. (2000). Learning a procedural task: effectiveness of multimedia presentations. *Applied Cognitive Psychology*, 14(6):555–575.

Molenberghs, P., Brander, C., Mattingley, J. B., and Cunnington, R. (2010). The role of the superior temporal sulcus and the mirror neuron system in imitation. *Human Brain Mapping*, 31(9):1316–1326.

Molenberghs, P., Cunnington, R., and Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews*, 36(1):341–349.

Montague, P., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N., and Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during Linked Social Interactions. *NeuroImage*, 16(4):1159–1164.

Mottolese, C., Richard, N., Harquel, S., Szathmari, A., Sirigu, A., and Desmurget, M. (2013). Mapping motor representations in the human cerebellum. *Brain*, 136(1):330–342.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology*, 20(8):750–756.

Nguyen, V. T., Sonkusare, S., Stadler, J., Hu, X., Breakspear, M., and Guo, C. C. (2017). Distinct Cerebellar Contributions to Cognitive-Perceptual Dynamics During Natural Viewing. *Cerebral Cortex*, 27(12):5652–5662.

Nissen, M. J. and Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology*, 19(1):1–32.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jaaskelainen, I. P., Hari, R., and Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, 109(24):9599–9604.

Nummenmaa, L., Lahnakoski, J. M., and Glerean, E. (2018). Sharing the social world via intersubject neural synchronisation. *Current Opinion in Psychology*, 24:7–14.

Nummenmaa, L., Smirnov, D., Lahnakoski, J. M., Glerean, E., Jaaskelainen, I. P., Sams, M., and Hari, R. (2014). Mental Action Simulation Synchronizes Action-Observation Circuits across Individuals. *Journal of Neuroscience*, 34(3):748–757.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97–113.

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., and Downing, P. E. (2010). Surface-Based Information Mapping Reveals Crossmodal Vision–Action Representations in Human Parietal and Occipitotemporal Cortex. *Journal of Neurophysiology*, 104(2):1077–1089.

Pajula, J., Kauppi, J.-P., and Tohka, J. (2012). Inter-Subject Correlation in fMRI: Method Validation against Stimulus-Model Based Analysis. *PLoS ONE*, 8(8):e41196.

Pajula, J. and Tohka, J. (2014). Effects of spatial smoothing on inter-subject correlation based analysis of FMRI. *Magnetic Resonance Imaging*, 32(9):1114–1124.

Pan, Y., Novembre, G., Song, B., Li, X., and Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *NeuroImage*, 183:280–290.

Penhune, V. B. and Steele, C. J. (2012). Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behavioural Brain Research*, 226(2):579–591.

Peterson, B. M., Wanders, I., Horne, K., Collier, S., Alexander, T., Kaspi, S., and Maoz, D. (1998). On Uncertainties in Cross-Correlation Lags and the Reality of Wavelength-dependent Continuum Lags in Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 110(748):660–670.

Piazza, E. A., Hasenfratz, L., Hasson, U., and Lew-Williams, C. (2018). Infant and adult brains are coupled to the dynamics of natural communication. *bioRxiv*.

Proville, R. D., Spolidoro, M., Guyon, N., Dugué, G. P., Selimi, F., Isope, P., Popa, D., and Léna, C. (2014). Cerebellum involvement in cortical sensorimotor circuits for the control of voluntary movements. *Nature Neuroscience*, 17(9):1233–1239.

Ramachandran, V. S. and Oberman, L. M. (2006). Broken mirrors. *Scientific American*, 295(5):62–69.

Raudies, F. (2013). Optic flow. *Scholarpedia*, 8(7):30724. revision #149632.

Regev, M., Honey, C. J., Simony, E., and Hasson, U. (2013). Selective and Invariant Neural Responses to Spoken and Written Narratives. *Journal of Neuroscience*, 33(40):15978–15988.

Reithler, J., van Mier, H. I., and Goebel, R. (2010). Continuous motor sequence learning: Cortical efficiency gains accompanied by striatal functional reorganization. *NeuroImage*, 52(1):263–276.

Rizzolatti, G. (2014). Imitation: mechanisms and importance for human culture. *Rendiconti Lincei*, 25(3):285–289.

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27(1):169–192.

Rizzolatti, G., Fabbri-Destro, M., and Cattaneo, L. (2009). Mirror neurons and their clinical relevance. *Nature Clinical Practice Neurology*, 5(1):24–34.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2(9):661–670.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology*, 12(2):149–154.

Rizzolatti, G. and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4):264–274.

Rizzolatti, G. and Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nature Reviews Neuroscience*, 17(12):757–765.

Ronconi, L., Casartelli, L., Carna, S., Molteni, M., Arrigoni, F., and Borgatti, R. (2016). When one is Enough: Impaired Multisensory Integration in Cerebellar Agenesis. *Cerebral Cortex*, page bhw049.

Rozzi, S., Ferrari, P. F., Bonini, L., Rizzolatti, G., and Fogassi, L. (2008). Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *European Journal of Neuroscience*, 28(8):1569–1588.

Sacheli, L. M., Verga, C., Arcangeli, E., Banfi, G., Tettamanti, M., and Paulesu, E. (2019). How Task Interactivity Shapes Action Observation. *Cerebral Cortex*, page 13.

Sakreida, K., Higuchi, S., Di Dio, C., Ziessler, M., Turgeon, M., Roberts, N., and Vogt, S. (2017). Cognitive Control Structures in the Imitation Learning of Spatial Sequences and Rhythms—An fMRI Study. *Cerebral Cortex*.

Schilbach, L. (2016). Towards a second-person neuropsychiatry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686):20150081.

Schippers, M. B., Gazzola, V., Goebel, R., and Keysers, C. (2009). Playing Charades in the fMRI: Are Mirror and/or Mentalizing Areas Involved in Gestural Communication? *PLoS ONE*, 4(8):e6801.

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences*, 107(20):9388–9393.

Schlerf, J. E., Verstynen, T. D., Ivry, R. B., and Spencer, R. M. C. (2010). Evidence of a Novel Somatopic Map in the Human Neocerebellum During Complex Actions. *Journal of Neurophysiology*, 103(6):3330–3336.

Schmahmann, J. D. (2019). The cerebellum and cognition. *Neuroscience Letters*, 688:62–75.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., and Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, 111(43):E4687–E4696.

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., and Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, 7:12141.

Spiegelhalder, K., Ohlendorf, S., Regen, W., Feige, B., Tebartz van Elst, L., Weiller, C., Hennig, J., Berger, M., and Tüscher, O. (2014). Interindividual synchronization of brain activity during live verbal communication. *Behavioural Brain Research*, 258:75–79.

Stansbury, D. E., Naselaris, T., and Gallant, J. L. (2013). Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex. *Neuron*, 79(5):1025–1034.

Steele, C. J. and Penhune, V. B. (2010). Specific Increases within Global Decreases: A Functional Magnetic Resonance Imaging Investigation of Five Days of Motor Sequence Learning. *Journal of Neuroscience*, 30(24):8332–8341.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32):14425–14430.

Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J.-M., Oostenveld, R., Hagoort, P., and Toni, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings of the National Academy of Sciences*, 111(51):18183–18188.

Stoodley, C. J., MacMore, J. P., Makris, N., Sherman, J. C., and Schmahmann, J. D. (2016). Location of lesion determines motor vs. cognitive consequences in patients with cerebellar stroke. page 11.

Stoodley, C. J. and Schmahmann, J. D. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage*, 44(2):489–501.

Stoodley, C. J., Valera, E. M., and Schmahmann, J. D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. *NeuroImage*, 59(2):1560–1570.

Tanabe, H. C., Kato, M., Miyauchi, S., Hayashi, S., and Yanagida, T. (2005). The sensorimotor transformation of cross-modal spatial information in the anterior intraparietal sulcus as revealed by functional MRI. *Cognitive Brain Research*, 22(3):385–396.

Thomas, R., De Sanctis, T., Gazzola, V., and Keysers, C. (2018). Where and how our brain represents the temporal structure of observed action. *NeuroImage*.

Tracy, J., Flanders, A., Madi, S., Laskas, J., Stoddard, E., Pyrros, A., Natale, P., and DelVecchio, N. (2003). Regional brain activation associated with different performance patterns during learning of a complex motor skill. *Cerebral Cortex*, 13(9):904–910.

Turella, L., Pierno, A., Tubaldi, F., and Castiello, U. (2009). Mirror neurons in humans: Consisting or confounding evidence? *Brain and Language*, 108(1):10–21.

Umilta, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., Gallese, V., and Rizzolatti, G. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences*, 105(6):2209–2213.

van Es, D. M., van der Zwaag, W., and Knapen, T. (2019). Topographic Maps of Visual Space in the Human Cerebellum. *Current Biology*, page S0960982219304105.

Van Overwalle, F. and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, 48(3):564–584.

Van Overwalle, F., Baetens, K., Mariën, P., and Vandekerckhove, M. (2014). Social cognition and the cerebellum: A meta-analysis of over 350 fMRI studies. *NeuroImage*, 86:554–572.

Vigneswaran, G., Philipp, R., Lemon, R. N., and Kraskov, A. (2013). M1 Corticospinal Mirror Neurons and Their Role in Movement Suppression during Action Observation. *Current Biology*, 23(3):236–243.

Vingerhoets, G., Nys, J., Honoré, P., Vandekerckhove, E., and Vandemaele, P. (2013). Human Left Ventral Premotor Cortex Mediates Matching of Hand Posture to Object Use. *PLoS ONE*, 8(7):e70480.

Vogt, S., Buccino, G., Wohlschläger, A. M., Canessa, N., Shah, N. J., Zilles, K., Eickhoff, S. B., Freund, H.-J., Rizzolatti, G., and Fink, G. R. (2007). Prefrontal involvement in imitation learning of hand actions: Effects of practice and expertise. *NeuroImage*, 37(4):1371–1383.

Wade, S. and Hammond, G. (2015). Anodal transcranial direct current stimulation over premotor cortex facilitates observational learning of a motor sequence. *European Journal of Neuroscience*, 41(12):1597–1602.

Walther, D. B., Caddigan, E., Fei-Fei, L., and Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience*, 29(34):10573–10581.

Wiestler, T. and Diedrichsen, J. (2013). Skill learning strengthens cortical representations of motor sequences. *eLife*, 2.

Wong, A., Marcus, N., Ayres, P., Smith, L., Cooper, G. A., Paas, F., and Sweller, J. (2009). Instructional animations can be superior to statics when learning human motor skills. *Computers in Human Behavior*, 25(2):339–347.

Zadbood, A., Chen, J., Leong, Y., Norman, K., and Hasson, U. (2017). How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cerebral Cortex*, 27(10):4988–5000.

A

Appendix

A.1 Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Stefan Glasauer, for offering me this great research project, supporting me throughout it and allowing me to follow my own ideas - and for letting me explore a vastness of topics ranging from behavioral science to cognitive science to neuroscience to data science and to diverse statistical methods.

Moreover, I would like to thank Dr. Virginia Flanagin not only for being a great collaborator but also for having helpful and supportive comments throughout the whole process of this thesis.

Special thanks got to my other TAC members, Prof. Michael Zehetleitner and Dr. Klaus Wunderlich for their support and helpful comments.

I would also like to thank the GSN for supporting me by giving me the opportunity to connect with many other PhD students, to attend great courses, go on excellent trips and for funding me to go to conferences.

During the course of this big experiment, I had support from several people, whom I am grateful to. I would like to thank Prof. Peter zu Eulenburg for his help with and enablement of the new multiband MRI sequences leading to the great data resolution of this thesis. Also I would like to thank Prof. Paul Taylor for helpful comments about design aspects of this experiment. I am very grateful to Hans Hintermaier for building the MRI setup: the cardboard folding table and the camera adapter for the MRI double mirror (and for awesome chats about surfing!). I am much very much obliged to Mathias Hübner for being the best right-hand man at the MRI: for being enthusiastic and very helpful when trying out and improving the setup, and for performing the scannings with me (and for nerdy chats about Star Trek!). Thank you so much, Mathias! Much thanks also go to Natalia Filvarova for helping with the subject recruitment and with the scannings. Last but not least, very special thanks go to Dr. Thomas 'Thommi' Stephan for being very helpful throughout the whole process of the experiment - from logistics to technical aspects - and especially for always bothering with my IT requests.

This thesis would of course have been impossible without my great colleagues.
I would like to give my special thanks to my colleagues Isabelle Garzorz, Nisha Dalal, Christopher Roppelt, Judita Huber, Alicia Costalago Meruelo, Dian Anggraini and Mehrdad Salmasi for their incredible social support throughout this thesis - may it be the positive, fun and encouraging atmosphere they created or the fact that they went with me through all the ups and downs of this work. I am especially grateful to Isabelle for being my last-minute corrector.

Very special thanks go to Alexander Knorr and Mehrdad Salmasi for being go-to persons for discussions about mathematical aspects and to Rainer Bögle for being a go-to person about fMRI questions (Rainer, I'll miss your meme-broadcast!). Moreover, I would like to express my gratitude to Theresa Raiser and Joshua Yudice for not only being fun colleagues but also for practicing my origami. I am especially grateful and indebted to Theresa for being such a patient and engaged instructor.
Moreover, I would like to thank all students and post-docs of the Forschungshaus for the fun time and great discussions I had, may it be around lunch time, coffee or ice cream breaks, or during the barbecues.

Apart from my work colleagues I would also like to express my greatest gratitude to my father, his wife and my family for their continuous support during the whole process and every aspect of this thesis.
I am also extremely grateful for my amazing friends who accompanied me throughout this journey.

Lastly, I want to express my greatest and endless gratitude to my partner Chris. For making a direct contribution to this thesis by being my guinea pig during the beginning of this experiment (bothering with all the 'can you try out if you can fold this blindly' requests or volunteering as a test subject in the MRI). But of course, in particular and especially, I would like to thank you, Chris, for your enormous unconditional support throughout absolutely every stage and aspect of this thesis.

A.2 Eidesstattliche Versicherung/Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation 'Human brain-to-brain synchrony in a naturalistic setting: an fMRI study on observational learning' selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation 'Human brain-to-brain synchrony in a naturalistic setting: an fMRI study on observational learning' is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München, 29.05.2019

Kathrin Kostorz

A.3 Author contributions

Both manuscripts (chapters 2 and 3)
K.K., V.L.F. and S.G. conceived of the experiment; K.K. designed and planned the experiment with support of V.L.F. and S.G.; K.K. and V.L.F. programmed the experiment; K.K. corrected the videos with support of S.G.; K.K. recruited and trained the subjects, performed the measurements and preprocessed the data.

First manuscript (chapter 2)
K.K. and S.G. analyzed the data with support of V.L.F; K.K. created the figures and wrote the manuscript with support of S.G. and V.L.F.

Second manuscript (chapter 3)
K.K. and S.G. analyzed the data; K.K. created the figures and wrote the manuscript with support of S.G. and V.L.F.